

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:43:52 ; Search time 289.144 Seconds
(without alignments)
10338.975 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caatgaaaaagttgttc.....ctcacagaccacaaagtacc 123

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_esti:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gssi:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	99.2	277	10	BF420446
2	122	99.2	292	9	AI598992
3	122	99.2	316	9	AI412434
4	122	99.2	404	10	BF549054

5	122	99.2	405	10	BE104696
6	122	99.2	407	9	AI009011
7	122	99.2	429	9	AI059212
8	122	99.2	455	9	AI012196
9	122	99.2	459	13	BQ780699
10	122	99.2	463	10	BF548903
11	122	99.2	485	9	AA899757
12	122	99.2	472	13	BQ190077
13	122	99.2	477	9	AW520555
14	122	99.2	489	9	AA859879
15	122	99.2	493	10	BF554834
16	122	99.2	522	14	C06842
17	122	99.2	565	12	BI273986
18	122	99.2	615	10	BG380448
19	122	99.2	639	10	BE111691
20	122	99.2	640	13	BQ780117
21	122	99.2	644	14	CA339407
22	122	99.2	687	12	BM389059
23	122	99.2	705	13	BQ200122
24	122	99.2	760	13	BQ191985
25	120.4	97.9	316	9	AI409795
26	120.4	97.9	557	10	BG665005
27	118.8	96.6	412	13	BO780657
28	118.8	96.6	517	9	AA859940
29	118.8	96.6	553	9	AA799542
30	118.8	96.6	636	13	BQ780658
31	118.8	96.6	638	12	BI283790
32	117.2	95.3	294	9	AI236740
33	117.2	95.3	469	12	BI276611
34	115.6	94.0	431	9	AI410870
35	115.6	94.0	446	12	BI286743
36	112.4	91.4	256	9	AI234716
37	110.8	90.1	427	9	AI178082
38	107.6	87.5	221	12	BM200395
39	107.6	87.5	225	10	BE136269
40	107.6	87.5	240	9	AW323013
41	107.6	87.5	289	12	BM200390
42	107.6	87.5	279	13	BU757687
43	107.6	87.5	283	9	AW122670
44	107.6	87.5	288	9	AV128839
45	107.6	87.5	294	14	W13019

ALIGNMENTS

RESULT 1	BF420446	277 bp	mRNA	linear	EST 28-NOV-2000
LOCUS	UI-R-BJ2-bpw-e-11-0-UI.s1	UI-R-BJ2	Rattus norvegicus	cdna clone	
DEFINITION	UI-R-BJ2-bpw-e-11-0-UI 3', mRNA sequence.				
ACCESSION	BF420446				
VERSION	BF420446.1	GI:11408475			
KEYWORDS	EST.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
REFERENCE	1 (bases 1 to 277)				
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery				
JOURNAL	Genome Res. 6 (9), 791-806 (1996)				
MEDLINE	97044477				
COMMENT	8889548				
	Contact: Soares, MB				
	Coordinated Laboratory for Computational Genomics				
	University of Iowa				
	375 Newton Road , 4156				
	MEBRF, Iowa City, IA 52242, USA				
	Tel: 319 335 8250				
	Fax: 319 335 9565				
	Email: bento-soares@uiowa.edu				

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized atrium at 16.5 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..277
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bpw-e-11-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ2"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at ratstat.eng.uiowa.edu. The subtraction has been previously described in (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG LIB=UI-R-BJ2
TAG_TISSUE=atrium at 16.5 dpc
TAG_SEQ=GATTC"

BASE COUNT 85 a 58 c 69 g 65 t
ORIGIN

Query Match 99.2%; Score 122; DB 10; Length 277;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTTCCCAAGTGAG 60
Db 66 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTTCCCAAGTGAG 125

QY 61 TTAGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120
Db 126 TTAGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 185

QY 121 AC 122
Db 186 AC 187

RESULT 2
AI598992
LOCUS
DEFINITION
EST250695 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
REMEG05.3' end, mRNA sequence.
ACCESSION
AI598992
VERSION
AI598992.1 GI:4608040
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
1 (bases 1 to 292)
AUTHORS
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat Gene Index
JOURNAL
Unpublished
CONTACT: Lee, NH
The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA
Tel.: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source

Location/Qualifiers
1..292
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="taxon:10118"
/clone="REMEG05"
/dev_stage="embryo 8, 12, 18 dpc"
/clone_lib="Normalized rat embryo, Bento Soares"
/note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 89 a 60 c 70 g 73 t
ORIGIN

Query Match 99.2%; Score 122; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 7.2e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTTCCCAAGTGAG 60
Db 97 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTTCCCAAGTGAG 156

QY 61 TTAGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120
Db 157 TTAGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 216

QY 121 AC 122
Db 217 AC 218

RESULT 3
AI412434
LOCUS
DEFINITION
EST240733 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBRD065.3' end, mRNA sequence.
ACCESSION
AI412434
VERSION
AI412434.1 GI:4255938
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
1 (bases 1 to 316)
AUTHORS
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat Gene Index
JOURNAL
Unpublished
CONTACT: Lee, NH
The Institute for Genomic Research

FEATURES
source

Location/Qualifiers
1..316
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="taxon:10118"
/clone="RBRD065"
/note="Organ: brain; Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 98 a 62 c 76 g 80 t
ORIGIN

Query Match 99.2%; Score 122; DB 9; Length 316;

Best Local Similarity 100.0%; Pred. No. 7.5e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTCTTAGTGGTGAAGGCCCAACACTGTGTTCCTCCAGTGAG 60
|||||
Db 97 CAATTGAAAGTTTCTTAGTGGTGAAGGCCCAACACTGTGTTCCTCCAGTGAG 156
|||||
QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 120
|||||
Db 157 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 216
|||||
QY 121 AC 122
||
Db 217 AC 218

RESULT 4
BF549054/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF549054 404 bp mRNA linear EST 11-DEC-2000
UI-R-A0-ag-b-01-0-UI.r1 UI-R-A0 Rattus norvegicus cDNA clone
UI-R-A0-ag-b-01-0-UI 5', mRNA sequence.
BF549054
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PubMed

1 (bases 1 to 404)
Bonaldio,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward

Location/Qualifiers
i. .404
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-ag-b-01-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-A0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."

BASE COUNT
ORIGIN
92 a 98 c 83 g 131 t
Query Match 99.2%; Score 122; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 8.3e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTCTTAGTGGTGAAGGCCCAACACTGTGTTCCTCCAGTGAG 60
|||||

Db 368 CAATTGAAAGTTTCTTAGTGGTGAAGGCCCAACACTGTGTTCCTCCAGTGAG 309
QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 120
|||||
Db 308 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 249
|||||
QY 121 AC 122
||
Db 248 AC 247

RESULT 5
BE104696
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE104696 405 bp mRNA linear EST 13-JUN-2000
UI-R-BX0-ars-e-01-0-UI.s1 UI-R-BX0 Rattus norvegicus cDNA clone
UI-R-BX0-ars-e-01-0-UI 3', mRNA sequence.
BE104696
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PubMed

1 (bases 1 to 405)
Bonaldio,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA sequence:
1-22, >at rich#Low complexity
Seq primer: M13 Forward
POLYA=Yes.

Location/Qualifiers
i. .405
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BX0-ars-e-01-0-UI"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BX0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BX0
library is derived from 13 dpc whole embryo tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
rategest.eng.uiowa.edu.
TAG_SEQ=None found"

BASE COUNT
ORIGIN
131 a 76 c 87 g 110 t 1 others
Query Match 99.2%; Score 122; DB 10; Length 405;
Best Local Similarity 100.0%; Pred. No. 8.3e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTCTTAGTGGTGAAGGCCCAACACTGTGTTCCTCCAGTGAG 60
|||||

```

Db      117 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 176
QY      61 TTAGTTGTACAGAACCGGTTAGCACTAGCGCTTGACAGAACCTCAGACCCCAAGGT 120
Db      177 TTAGTTGTACAGAACCGGTTAGCACTAGCGCTTGACAGAACCTCAGACCCCAAGGT 236
QY      121 AC 122
Db      237 AC 238

RESULT 6
LOCUS   AI009011
DEFINITION EST203462 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION AI009011
VERSION   AI009011.1 GI:3222843
KEYWORDS EST.
SOURCE   Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 407)
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat
Gene Index
Unpublished
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..407
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2020087"
/db_xref="taxon:10118"
/clone="REMBH20"
/dev_stage="embryo 8, 12, 18 dpc"
/clone_lib="Normalized rat embryo, Bento Soares"
/vector="PT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 135 a 79 c 97 g 96 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 407;
Best Local Similarity 100.0%; Pred. No. 8.3e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 91 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 150
QY 61 TTAGTTGTACAGAACCGGTTAGCACTAGCGCTTGACAGAACCTCAGACCCCAAGGT 120
Db 151 TTAGTTGTACAGAACCGGTTAGCACTAGCGCTTGACAGAACCTCAGACCCCAAGGT 210
QY 121 AC 122
Db 211 AC 212

RESULT 7
LOCUS   AI059212
DEFINITION UI-R-C1-1b-e-11-0-UI-s1 UI-R-C1 Rattus norvegicus cDNA clone
ACCESSION AI059212
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 429)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult lung library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1..429
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C1-1b-e-11-0-UI"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-C1"
/vector="PT73Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C1
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The UI-R-E1 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dr track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C1) was constructed as follows: PCR
amplified cDNA inserts from UI-R-C0 clones from which 3'
ESTs had been derived was used as a driver in a
hybridization with the UI-R-C0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-C1
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."
BASE COUNT 130 a 87 c 91 g 121 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 429;
Best Local Similarity 100.0%; Pred. No. 8.5e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 146 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 205

```



```

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAAGGT 120
|||
Db 206 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAAGGT 265
|||
QY 121 AC 122
||
Db 266 AC 267

RESULT 8
AI012196 455 bp mRNA linear EST 15-JUN-1998
DEFINITION EST206647 Normalized rat placenta, Bento Soares Rattus sp. CDNA
clone RPLAT85 3' end, mRNA sequence.
ACCESSION AI012196
VERSION AI012196.1 GI:3226028
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 455)
AUTHORS Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL Unpublished
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: rhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..455
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="taxon:10118"
/clone="RPLAT85"
/clone_lib="Normalized rat placenta, Bento Soares"
/notes="Organ: Placenta; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 156 a 92 c 102 g 105 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 8.8e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTGAG 60
|||
Db 91 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTGAG 150
|||
QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAAGGT 120
|||
Db 151 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAAGGT 210
|||
QY 121 AC 122
||
Db 211 AC 212

RESULT 9
BQ780699 459 bp mRNA linear EST 26-JUL-2002
LOCUS BQ780699
DEFINITION UI-R-FF0-cpc-k-22-0-UI.s1 UI-R-FF0 Rattus norvegicus CDNA clone
ACCESSION BQ780699
VERSION BQ780699.1 GI:21989171
KEYWORDS EST.

```

```

SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 459)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1..459
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="UI-R-FF0-cpc-k-22-0-UI"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FF0"
/notes="Vector: pT7T3-Pac (Pharmacia) with a modified
polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-FF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS tumor line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTAATGGACG,
CAATCTTGTA.
TAG_LIB=UI-R-FF0
TAG_TISSUE=cartilage
TAG_SEQ=CTAATGGACG"
BASE COUNT 145 a 91 c 103 g 120 t
ORIGIN
Query Match 99.2%; Score 122; DB 13; Length 459;
Best Local Similarity 100.0%; Pred. No. 8.8e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTGAG 60
|||
Db 114 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTGAG 173
|||
QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAAGGT 120
|||
Db 174 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAAGGT 233
|||
QY 121 AC 122
||
Db 234 AC 235

RESULT 10
BF548903/c

```

LOCUS
 DEFINITION BF548903 463 bp mRNA linear EST 11-DEC-2000
 UI-R-A0-ag-b-08-0-UI.r1 UI-R-A0 Rattus norvegicus cDNA clone
 UI-R-A0-ag-b-08-0-UI 5', mRNA sequence.

ACCESSION
 VERSION BF548903
 KEYWORDS BF548903.1 GI:11640010
 SOURCE EST.
 ORGANISM Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (Bases 1 to 463)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)

JOURNAL
 MEDLINE 9704477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 cDNA Library Preparation: M.B. Soares Lab Clone distribution:
 clones will be available through Research Genetics (www.resgen.com)
 This clone is also available through the I.M.A.G.E. Consortium at
 LLNL (info@image.llnl.gov). IMAGE ID= 1791813
 Seq primer: M13 Forward.

FEATURES
 source
 1..463
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-A0-ag-b-08-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-A0"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; This library
 consists of a mixture of individually tagged normalized
 libraries constructed from rat placenta, adult lung, brain
 liver, kidney, heart, spleen, ovary, and muscle. The tag
 is a string of 3-5 nucleotides present between the Not I
 site and the oligo-dT track which allows identification of
 the library of origin of a clone within the mixture."
 BASE COUNT 109 a 106 c 93 g 155 t

Query Match 99.2%; Score 122; DB 10; Length 463;
 Best Local Similarity 100.0%; Pred. No. 8.8e-31;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCAGTGAG 60
 |||
 Db 368 CAATTGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCAGTGAG 309
 |||

QY 61 TTAGTTGTACAGACGGCGTTAGCACTAGCGTTGACAGAACCTCAGACACCAAGGT 120
 |||
 Db 308 TTAGTTGTACAGACGGCGTTAGCACTAGCGTTGACAGAACCTCAGACACCAAGGT 249
 |||

QY 121 AC 122
 ||
 Db 248 AC 247

RESULT 11
 AA899757
 LOCUS
 DEFINITION UI-R-E0-da-b-11-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
 UI-R-E0-da-b-11-0-UI 3' similar to gi|464185|dbj|D25274|HUMPO2ST9

LOCUS
 DEFINITION AA899757 472 bp mRNA linear EST 30-APR-2002
 UI-R-CN1-cjs-h-07-0-UI.s3 UI-R-CN1 Rattus norvegicus cDNA clone
 UI-R-CN1-cjs-h-07-0-UI 3', mRNA sequence.

ACCESSION
 VERSION AA899757.1 GI:4232251
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (Bases 1 to 465)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)

JOURNAL
 MEDLINE 9704477
 PUBMED 8889548
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3035111.
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
 distribution: clones will be available through Research Genetics
 Seq primer: M13 Forward.

FEATURES
 source
 1..465
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-E0-da-b-11-0-UI"
 /dev_stage="embryonic"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-E0"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: NotI; Site 2: EcoRI; This library
 consists of a mixture of individually tagged normalized
 libraries constructed from 8, 12 and 18-day embryo. The
 tag is a string of 3-5 nucleotides present between the
 Not I site and the oligo-dT track which allows
 identification of the library of origin of a clone within
 the mixture."
 BASE COUNT 155 a 93 c 99 g 118 t

Query Match 99.2%; Score 122; DB 9; Length 465;
 Best Local Similarity 100.0%; Pred. No. 8.8e-31;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCAGTGAG 60
 |||
 Db 112 CAATTGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCAGTGAG 171
 |||

QY 61 TTAGTTGTACAGACGGCGTTAGCACTAGCGTTGACAGAACCTCAGACACCAAGGT 120
 |||
 Db 172 TTAGTTGTACAGACGGCGTTAGCACTAGCGTTGACAGAACCTCAGACACCAAGGT 231
 |||

QY 121 AC 122
 ||
 Db 232 AC 233

RESULT 12
 BQ190077
 LOCUS
 DEFINITION UI-R-CN1-cjs-h-07-0-UI.s3 UI-R-CN1 Rattus norvegicus cDNA clone
 UI-R-CN1-cjs-h-07-0-UI 3', mRNA sequence.

ACCESSION
 VERSION BQ190077.1 GI:20365628
 KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 472)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
REFERENCE Normalization and subtraction: two approaches to facilitate gene
AUTHORS discovery
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized rat eye library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.regen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
1. .472
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="rRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CNI-GJs-h-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CNI"
/note="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CNI
library is a subtracted library derived from the following
pool of seven normalized rat libraries: normalized rat
seminal vesicles, normalized rat penis, normalized rat
bladder, normalized rat cervix, normalized rat brown
adipose, normalized rat fundus, and normalized rat
salivary gland. It was constructed according to the
procedure described by Bonaldo, Lennon & Soares (Genome
Research Volume 6: 791-806, 1996). For construction of
the CNI library, plasmid DNA from the pool of normalized
libraries was electroporated into competent bacteria for
the production of single-stranded circular DNA. This was
then used as a tracer in a subtractive hybridization with
a driver (PCR amplified inserts from a plasmid DNA template
preparation) comprising: a) a pool of about 34,000 clones
from the Rat Unigene Set corresponding to plates R-5-AR-NN
excluding plates R-5-AM and MN. This pool represented 40%
of the final driver population. b) a pool of about 29,000
clones from subtracted libraries CA0 and CA1 corresponding
to plates R-CA0-AWV through R-CA0-AXS, R-CA0-AZX through
R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS,
R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA,
R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through
R-CA0-BOU, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through
R-CA1-BDA, R-CA1-BHZ through R-CA1-BJF, R-CA1-BJR,
R-CA1-BJT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF,
R-CA1-BKI, R-CA1-BKT, R-CA1-BLF, R-CA1-BLH through
R-CA1-BLN, R-CA1-BLS, R-CA1-BLU-V, R-CA1-BNR, and
R-CA1-BUE. The resulting pool represented 20% of the
final driver population. c) a pool of about 15,000 clones
from non-normalized libraries CS0s, CT0s, CW0s, CX0s
and normalized libraries CS0, CT0, CU0, CW0, and CX0

corresponding to plates R-CS0s-CBD through R-CS0s-CBO,
R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through
R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN
through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV
R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN,
R-CU0-BUQ through R-CU0-BVL, R-CW0-BVV through R-CW0-BWP,
R-CX0-BXN through R-CW0-BXO, R-CX0-BWQ through R-CX0-BXM.
The resulting pool represented 5% of the final driver
population. d) a pool of about 5,000 clones (1,000 from
non-normalized eye library CV0 and 4,000 from normalized
eye library CV1) corresponding to plates R-CV0-BRH through
R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through
R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool
represented about 5% of the final driver population. e) A
pool of about 10,000 clones from subtracted library BS2,
BV0 and BV0p (7-9.5 kb cDNA library fraction from rat
whole embryo), and BX0 (0.5-7kb cDNA library fraction from
rat whole embryo) corresponding to plates R-BS2-BDB
through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0p-AOI
through R-BV0p-AOX, and R-BX0-AQY through R-BX0-ASH. The
resulting pool represented 5% of the final driver
population. f) a pool of about 7,000 clones from the
seven non-normalized libraries that make up the tracer:
including CV0, CZ0, DA0, DB0, DCO, DD0, and DE0
corresponding to plates R-CY0-BXP through R-CY0-BXZ,
R-CZ0-BXA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ
through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ
through R-DB0-BZA, R-DC0-BZI through R-DC0-BZO, R-DC0-CAY
through R-DC0-CBA, R-DD0-BZR through R-DD0-CAA, The
R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The
resulting pool represented about 10% of the final driver
population. g) a pool of about 2,000 clones from the pool
of normalized libraries, CNO, that makes up the tracer.
The corresponding plates are R-CNO-BKM through R-CNO-BLD,
R-CNO-BLG, R-CNO-BLP through R-CNO-BLR, R-CNO-BLT,
R-CNO-BLW-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML.
This pool represented 5% of the final driver population.
h) a pool of the 28 most abundant clones in the CNO pool
corresponding to the following addresses: bkx-a-09-0-UI,
bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI,
bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI,
bkx-a-05-0-UI, bkz-a-06-0-UI, bkz-a-11-0-UI, bkz-c-06-0-UI,
bkz-c-09-0-UI, bkz-d-10-0-UI, bla-a-01-0-UI,
bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI,
blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI,
blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI,
blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5%
of the final driver population. i) One abundant CNO clone
(corresponding to the address bkz-a-11-0-UI) was digested
with Not I and Eco RI and the resulting insert was gel
purified. This purified insert was added directly to the
driver so that it represented 5% of the final driver
population.

TAG LI8-UI-R-CNI
TAG-TISSUE-rat eye
TAG-SEQ-CAGCC"

BASE COUNT 154 a 102 g 123 t
ORIGIN

Query Match 99.2%; Score 122; DB 13; Length 472;
Best Local Similarity 100.0%; Pred. No. 8.9e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTGTCTAGTGGTCCGAAGGCCCAACTGTGTCTTCCAGTCTAG 60
Db 113 CAATTGAAAAAGTTTGTCTAGTGGTCCGAAGGCCCAACTGTGTCTTCCAGTCTAG 172
QY 61 TTAGGTTTACAGAACGCGGTTAGCCTTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 120
Db 173 TTAGGTTTACAGAACGCGGTTAGCCTTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 232
QY 121 AC 122

```

Db          233 AC 234

RESULT 13
AW520555
LOCUS
DEFINITION UI-R-BJ0p-afx-g-09-0-UI.s1 477 bp mRNA linear EST 06-MAR-2000
            UI-R-BJ0p-afx-g-09-0-UI 3', mRNA sequence.
ACCESSION AW520555
VERSION    AW520555.1 GI:7162933
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 477)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized Av canal at 15 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.

FEATURES             Location/Qualifiers
     source            1..477
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ0p-afx-g-09-0-UI"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-BJ0p"
                     /note="Vector: pTT73D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0p
                     library is a subtracted library derived from the UI-R-AAL,
                     UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
                     UI-R-AG1 libraries. These libraries represent tissues from
                     rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
                     at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
                     canal at 15 dpc, and ventricle at 13 dpc. The tag is a
                     string of 5-6 nucleotides present between the Not I site
                     and the oligo-dT track. The library was constructed as
                     described by Bonaldo, Lennon and Soares, Genome Research
                     6: 791-806, 1996.
                     TAG_LIB=UI-R-BJ0p
                     TAG_TISSUE=AV canal at 15 dpc
                     TAG_SEQ=GAAGG"
BASE COUNT      157 a      103 g      124 t
ORIGIN
Query Match      99.2%; Score 122; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 8,9e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db          233 AC 234

RESULT 14
AW59879
LOCUS
DEFINITION UI-R-E0-cc-c-06-0-UI.s1 489 bp mRNA linear EST 03-JUL-1999
            UI-R-E0-cc-c-06-0-UI 3', similar to dbj|D25274|HUMPO2ST9 Human
            randomly sequenced mRNA, mRNA sequence.
ACCESSION AW59879
VERSION    AW59879.1 GI:4230416
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 489)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   On Mar 10, 1998 this sequence version replaced gi:2949399.
            Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics This clone is also available through the
            I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
            ID=1770064
            Seq primer: M13 Forward
            POLYA=No.

FEATURES             Location/Qualifiers
     source            1..489
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-E0-cc-c-06-0-UI"
                     /dev_stage="embryonic"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-E0"
                     /note="Vector: pTT73D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: NotI; Site 2: EcoRI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from 8, 12 and 18-day embryo. The
                     tag is a string of 3-5 nucleotides present between the
                     Not I site and the oligo-dT track which allows
                     identification of the library of origin of a clone within
                     the mixture."
BASE COUNT      158 a      96 c      107 g      128 t
ORIGIN
Query Match      99.2%; Score 122; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 9e-31;

```

Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 60
 |||||
 Db 112 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 171
 |||||

QY 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 120
 |||||
 Db 172 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 231
 |||||

QY 121 AC 122
 ||
 Db 232 AC 233

RESULT 15
 BF554834/c
 LOCUS
 DEFINITION UI-R-E0-cc-c-06-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
 UI-R-E0-cc-c-06-0-UI 5', mRNA sequence.

ACCESSION BF554834
 VERSION BF554834.1 GI:11664606
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 493)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT

Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 cDNA Library Preparation: M.B. Soares Lab Clone distribution:
 clones will be available through Research Genetics (www.resgen.com)
 This clone is also available through the I.M.A.G.E. Consortium at
 LLNL (info@image.llnl.gov). IMAGE ID= 1770664 The following
 repetitive elements were found in this cDNA sequence: 463-485,
 >AT rich#Low complexity
 Seq primer: M13 Forward.

FEATURES

Location/Qualifiers

1..493
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-E0-cc-c-06-0-UI"
 /dev_stage="embryonic"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-E0"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: NotI; Site 2: EcoRI; This library
 consists of a mixture of individually tagged normalized
 libraries constructed from 8, 12 and 18-day embryo. The
 tag is a string of 3-5 nucleotides present between the
 Not I site and the oligo-dT track which allows
 identification of the library of origin of a clone within
 the mixture."

BASE COUNT 130 a 109 c 95 g 159 t

ORIGIN

Query Match 99.2%; Score 122; DB 10; Length 493;
 Best Local Similarity 100.0%; Pred.No. 9.1e-31;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 60
 |||||
 Db 368 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 309
 |||||

QY 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 120
 |||||
 Db 308 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 249
 |||||

QY 121 AC 122
 ||
 Db 248 AC 247

Search completed: November 23, 2003, 15:49:53
 Job time : 297.144 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:17:01 ; Search time 35.0599 Seconds
(without alignments)
9470.385 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caatgaaaaaagtgtgttc.....ctcacagaccacaaaggtacc 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03:*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	123	100.0	123	Human rac1 genomic
2	123	100.0	348	Human rac1 gene re
3	123	100.0	1266	Human rac1 contig
4	106.6	86.7	422	Mouse Rab2 nucleot
5	96.6	78.5	447	Mouse Ecol103 nucle
6	79.6	64.7	352	Human rac1 gene re
7	79.6	64.7	1232	Human cDNA differe
8	79.6	64.7	1232	Gene #1632 used to

C 9	79.6	64.7	1318	23	AAS65569	DNA encoding novel
C 10	79.6	64.7	2051	21	AAF18236	Lung cancer associ
C 11	79.6	64.7	28567	25	ABT17030	Human MP21 gene Ra
C 12	57	46.3	353	25	ABX46267	Bovine EST associa
C 13	38.2	31.1	3740	22	AL07296	Human reproductive
C 14	34	27.6	50	24	ABZ04062	Human leukocyte ge
C 15	31.4	25.5	605	21	AA16478	Human colon cancer
C 16	31.4	25.5	605	24	ABU38065	Human colon tumour
C 17	28.8	23.4	202001	24	ABS52506	Human transporter
C 18	27.4	22.3	412	24	ABN94214	Gene #712 used to
C 19	27.4	22.3	412	24	ABL62818	Breast cancer rela
C 20	27.4	22.3	412	24	ABL63018	Breast cancer rela
C 21	27.4	22.3	412	24	ABL63234	Breast cancer rela
C 22	27.4	22.3	412	24	ABU80240	Human ovarian can
C 23	27.4	22.3	1089	21	AAC50142	Arabidopsis thalia
C 24	27.4	22.3	1091	21	AAC40517	Arabidopsis thalia
C 25	27.4	22.3	1658	24	ABT10169	Human breast cance
C 26	27.4	22.3	1689	22	AAF59609	Human cell cycle a
C 27	27.2	22.1	9965	24	ABL33526	Human immune syste
C 28	27	22.0	363	24	AB199532	Mouse ischaemic co
C 29	27	22.0	6693	25	ACA04014	cDNA downregulated
C 30	26.8	21.8	396	25	ABX05806	S. pneumoniae type
C 31	26.8	21.8	1617	24	ABK63792	Rat sequence diffe
C 32	26.8	21.8	6827	19	AAV52193	Streptococcus pneu
C 33	26.8	21.8	2162598	25	ABS56454	Streptococcus pneu
C 34	26.6	21.6	1030	21	AAC34296	Arabidopsis thalia
C 35	26.6	21.6	81001	22	AAF30035	Human apolipoprote
C 36	26.4	21.5	279	24	ABL60574	Human monocyte che
C 37	26.4	21.5	510	25	ABX74789	Human cDNA sequenc
C 38	26.4	21.5	1035	21	AA51499	Human growth facto
C 39	26.4	21.5	1035	24	ABS68613	DNA representing a
C 40	26.4	21.5	1035	25	ABX93147	Degenerate DNA enc
C 41	26.4	21.5	534720	19	AAV30458	Rhizobium species
C 42	26.4	21.5	536165	19	AAV30459	Rhizobium species
C 43	26.2	21.3	903	24	ABN70636	Streptococcus poly
C 44	26.2	21.3	2091	18	AAV25107	H. pylori cell env
C 45	26.2	21.3	2091	20	AAV75826	H. pylori outer me

ALIGNMENTS

RESULT 1
AAH22395
ID AAH22395 standard; DNA; 123 BP.
XX
AC AAH22395;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human rac1 genomic fragment probable 3' UTR SEQ ID NO:1.
XX
KW Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX
OS Homo sapiens.
XX
PN WO200138579-A2.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US32049.
XX
PR 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
PA (CURA-) CURAGEN CORP.
PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
DR WPI; 2001-355948/37.
XX

PT Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels

XX Disclosure; Page 7; 76pp; English.

XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in a reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX SQ Sequence 123 BP; 36 A; 28 C; 30 G; 29 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 123;
 Best Local Similarity 100.0%; Pred. No. 1.28-35;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 60
 DB 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 60

QY 61 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120
 DB 61 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120

QY 121 ACC 123
 DB 121 ACC 123

RESULT 2
 AAH22399/c
 ID AAH22399 standard; DNA; 348 BP.
 XX AC AAH22399;
 XX DT 22-AUG-2001 (first entry)
 XX DE Human rac1 gene related nucleotide sequence #3.
 XX KW Identification; toxic; hepatotoxic; differential gene expression;
 XX KW NSAID; non-steroidal antiinflammatory drug; ds.
 XX OS Homo sapiens.

XX PN WO200138579-A2.
 XX PD 31-MAY-2001.
 XX PF 21-NOV-2000; 2000WO-US32049.
 XX PR 22-NOV-1999; 99US-0166923.
 XX PR 18-FEB-2000; 2000US-0183531.
 XX PR 20-NOV-2000; 2000US-0717321.
 XX PA (CURA-) CURAGEN CORP.

XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
 XX DR WPI; 2001-355948/37.
 XX XX

PT Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels

XX Disclosure; Page 8-9; 76pp; English.

XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in a reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX SQ Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 348;
 Best Local Similarity 100.0%; Pred. No. 1.78-35;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 60
 DB 212 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 153

QY 61 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120
 DB 152 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 93

QY 121 ACC 123
 DB 92 ACC 90

RESULT 3
 AAH22396
 ID AAH22396 standard; DNA; 1266 BP.

XX AC AAH22396;
 XX DT 22-AUG-2001 (first entry)
 XX DE Human rac1 contig SEQ ID NO:2.

XX KW Identification; toxic; hepatotoxic; differential gene expression;
 XX KW NSAID; non-steroidal antiinflammatory drug; ds.

XX OS Homo sapiens.
 XX PN WO200138579-A2.
 XX PD 31-MAY-2001.
 XX PF 21-NOV-2000; 2000WO-US32049.
 XX PR 22-NOV-1999; 99US-0166923.
 XX PR 18-FEB-2000; 2000US-0183531.
 XX PR 20-NOV-2000; 2000US-0717321.
 XX PA (CURA-) CURAGEN CORP.

XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
 XX DR WPI; 2001-355948/37.
 XX XX

PT Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
PS Disclosure; Page 7; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in a reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
SQ Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 1266;
Best Local Similarity 100.0%; Pred. No. 2.8e-35;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTTCCAGTGAG 60
Db 146 CAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTTCCAGTGAG 205

QY 61 TTAGTTGTACAGAACGGCGTTAGCCTAGCGTTGACAGAACCTCAGACCCAAAGGT 120
Db 206 TTAGTTGTACAGAACGGCGTTAGCCTAGCGTTGACAGAACCTCAGACCCAAAGGT 265

QY 121 ACC 123
Db 266 ACC 268

RESULT 4
AAA9694/c
ID AAA9694 standard; cDNA; 422 BP.
XX
AC AAA9694;
XX
DT 08-JAN-2001 (first entry)
XX
XX Mouse Rab2 nucleotide sequence #1.
XX
XX Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
KW anti-allergic; antiasthmatic; nootropic; neuroprotective; anticonvulsant;
KW vulnary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
XX digestion disorder; wound healing disorder; gene therapy; ss.
OS Mus sp.
XX
XX WO200043419-A2.
XX
XX 27-JUL-2000.
XX
XX 20-JAN-2000; 2000WO-US01431.
XX
XX 20-JAN-1999; 99US-0116534.
PR 26-JAN-1999; 99US-0117274.
PR 26-JAN-1999; 99US-0117308.
PR 26-JAN-1999; 99US-0117309.
PR 26-JAN-1999; 99US-0117312.
PR 01-FEB-1999; 99US-0118177.
PR 01-FEB-1999; 99US-0118178.

PR 01-FEB-1999; 99US-0118179.
PR 09-FEB-1999; 99US-0119286.
PR 11-FEB-1999; 99US-0119998.
PR 11-FEB-1999; 99US-0119759.
XX
XX (RIGE-) RIGEL PHARM INC.
XX
XX Luo Y;
PI
XX WPI; 2000-482908/42.
DR
XX New nucleic acids encoding Exo proteins which are useful in the
PT diagnosis, treatment or prevention of exocytosis-mediated disorders
PT such as asthma, inflammation and allergies -
XX
XX Disclosure; Page 271-272; 305pp; English.
PS
XX The present sequence encodes a polypeptide which is associated with
CC the exocytosis pathway. cDNA molecules encoding proteins involved in
CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
CC screening. Novel proteins, termed Exo proteins, have been identified that
CC interact with known exocytosis-associated proteins such as GS27, alpha
CC snap, unc18-1, vamp3, snap-23, and the rab family of proteins.
CC Exo proteins and their agonists and antagonists are useful in the
CC diagnosis, treatment or prevention of exocytosis-mediated disorders
CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC diabetes, digestion disorders and wound healing disorders.
CC The nucleic acids, antagonists or agonists of Exo proteins are useful
CC in gene therapy. The nucleic acids are also useful for generating
CC transgenic or knock-out animals which can be used in the
CC development and screening of therapeutically useful reagents.
XX
SQ Sequence 422 BP; 95 A; 92 C; 83 G; 148 T; 4 other;

Query Match 86.7%; Score 106.6; DB 21; Length 422;
Best Local Similarity 91.8%; Pred. No. 2.2e-29;
Matches 112; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTTCCAGTGAG 60
Db 378 CAATTGAAAAAGTTTGTCTAGTGGTTGAAGGCCCAACACTGTGTTCCAGTGAG 319

QY 61 TTAGTTGTACAGAACGGCGTTAGCCTAGCGTTGACAGAACCTCAGACCCAAAGGT 120
Db 318 TTAGTTGTACAGAACGGCGTTAGCCTAGCGTTGACAGAACCTCAGACCCAAAGGT 259

QY 121 AC 122
Db 258 AC 257

RESULT 5
AAA9693/c
ID AAA9693 standard; cDNA; 447 BP.
XX
AC AAA9693;
XX
XX 08-JAN-2001 (first entry)
XX
XX Mouse Exo103 nucleotide sequence.
XX
XX Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
KW anti-allergic; antiasthmatic; nootropic; neuroprotective; anticonvulsant;
KW vulnary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
XX digestion disorder; wound healing disorder; gene therapy; ss.
OS Mus sp.
XX
XX WO200043419-A2.
XX
XX 27-JUL-2000.

XX PF 20-JAN-2000; 2000WO-US01431.
XX PR 20-JAN-1999; 99US-0116534.
XX PR 26-JAN-1999; 99US-0117274.
XX PR 26-JAN-1999; 99US-0117308.
XX PR 26-JAN-1999; 99US-0117309.
XX PR 26-JAN-1999; 99US-0117312.
XX PR 01-FEB-1999; 99US-0118177.
XX PR 01-FEB-1999; 99US-0118178.
XX PR 01-FEB-1999; 99US-0118179.
XX PR 09-FEB-1999; 99US-0119286.
XX PR 11-FEB-1999; 99US-0119998.
XX PR 11-FEB-1999; 99US-0119759.
XX PA (RIGE-) RIGEL PHARM INC.
XX PI Luo Y;
XX DR WPI; 2000-482908/42.
XX XX New nucleic acids encoding Exo proteins which are useful in the
PT diagnosis, treatment or prevention of exocytosis-mediated disorders
PT such as asthma, inflammation and allergies -
XX PS Disclosure; Page 271; 305pp; English.
XX CC The present sequence encodes a polypeptide which is associated with
CC the exocytosis pathway. cDNA molecules encoding proteins involved in
CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
CC screening. Novel proteins, termed Exo proteins, have been identified that
CC interact with known exocytosis-associated proteins such as GS27, alpha
CC snap, unc18-1, vamp3, snap-23, and the rab family of proteins.
CC Exo proteins and their agonists and antagonists are useful in the
CC diagnosis, treatment or prevention of exocytosis-mediated disorders
CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC diabetes, digestion disorders and wound healing disorders.
CC The nucleic acids, antagonists or agonists of Exo proteins are useful
CC in gene therapy. The nucleic acids are also useful for generating
CC transgenic or knock-out animals which can be used in the
CC development and screening of therapeutically useful reagents.
XX SQ Sequence 447 BP; 102 A; 95 C; 93 G; 157 T; 0 other;
Query Match 78.5%; Score 96.6; DB 21; Length 447;
Best Local Similarity 91.9%; Pred. No. 1.2e-25;
Matches 113; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
QY 1 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 59
DB 379 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 320
QY 60 GTTAGTTGTACAGACGGGTTAGCACTAGCGCTTGAAGAACCTTCACAGACCCCAAGG 119
DB 319 GTTAGTTGTACAGACGGGTTAGCACTAGCGCTTGAAGAACCTTCACAGACCCCAAGG 260
QY 120 TAC 122
DB 259 AAC 257
RESULT 6
AAH22400/c
ID AAH22400 standard; DNA; 352 BP.
XX AC AAH22400;
XX XX
DT 22-AUG-2001 (first entry)
XX DE Human rac1 gene related nucleotide sequence #4.
XX XX Identification; toxic; hepatotoxic; differential gene expression;
KW

KW NSAID; non-steroidal antiinflammatory drug; ds.
XX OS Homo sapiens.
XX PN WO200138579-A2.
XX PD 31-MAY-2001.
XX PF 21-NOV-2000; 2000WO-US32049.
XX PR 22-NOV-1999; 99US-0166923.
XX PR 18-FEB-2000; 2000US-0183531.
XX PR 20-NOV-2000; 2000US-0717321.
XX PA (CURA-) CURAGEN CORP.
XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX DR WPI; 2001-355948/37.
XX XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX PS Disclosure; Page 8-9; 76pp; English.
XX CC The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX SQ Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other;
Query Match 64.7%; Score 79.6; DB 22; Length 352;
Best Local Similarity 83.1%; Pred. No. 2.1e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 60
DB 216 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 157
QY 61 TTAGTTGTACAGACGGGTTAGCACTAGCGG--TTGACAGAACCTTCACAGACCCCAAG 118
DB 156 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACCTTCACAGACCCCAAG 97
QY 119 GTAC 122
DB 96 GAAC 93
RESULT 7
ABX83623/c
ID ABX83623 standard; cDNA; 1232 BP.
XX AC ABX83623;
XX XX
DT 14-AUG-2002 (first entry)
XX DE Human cDNA differentially expressed in granulocytic cells #194.
XX XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW

KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; peridontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX Homo sapiens.
XX WO200228999-A2.
XX 11-APR-2002.
XX 03-OCT-2001; 2001WO-US30821.
XX 03-OCT-2000; 2000US-237189P.
XX (GENE-) GENE LOGIC INC.
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX Detecting granulocyte activation by detecting differential expression
XX of genes associated with granulocyte activation, which serves as
XX diagnostic markers that is useful for monitoring disease states and
XX drug toxicity -
XX Claim 1; SEQ ID NO 194; 114pp; English.
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GA by contacting GC with an agent
XX that alters the expression of at least one gene in Gs; (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression in a sample of the tissue of gene(s) from Gs, where
XX the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX an allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX modulating GA; M3 is useful for screening an agent capable of modulating
XX GCA preferably in an inflammation in a tissue; M4 is useful for
XX detecting an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
XX reperfusion injury, ARDS, adult respiratory distress syndrome,
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,
XX peridontal disease; also bacterial infection, viral infection,
XX parasitic infection, protozoal infection, fungal infection and M5 is
XX useful for treating one of the above conditions. The present
XX sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;

Query Match 64.7%; Score 79.6; DB 24; Length 1232;
Best Local Similarity 83.1%; Pred. No. 3.4e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAAGTTTCTTAGTGTGCAAGGCCAACACACTGTCTTCCAGTGAG 60
DB 1135 CAATTAAAAAATTCGTTAGTGGCTGAAGGTCGCCACGCTGTATTCTCGCCAGTGAG 1076
QY 61 TTAGTGTGTACAGACGGGTTAGCACTAGCGC--TTGACAGAACCTCAGACCCCAAG 118
DB 1075 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGTTTACAGAACCTCAGACCCCAAG 1016
QY 119 GTAC 122
DB 1015 GAAC 1012
RESULT 8
ABN95134/C
ID ABN95134 standard; DNA; 1232 BP.
XX AC ABN95134;
XX DT 13-AUG-2002 (first entry)
XX DE Gene #1632 used to diagnose liver cancer.
XX KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX KW metastatic liver tumour; cytostatic; expression profile; disease state;
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US30589.
XX PR 02-OCT-2000; 2000US-237054P.
XX PA (GENE-) GENE LOGIC INC.
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX DR Diagnosing and detecting the progression of liver cancer,
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,
XX PT involves detecting the level of expression of two or more genes in a
XX liver tissue sample -
XX PS Claim 1; SEQ ID NO 1632; 298pp; English.
XX CC The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumor in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN9503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;

Query Match 64.7%; Score 79.6; DB 24; Length 1232;
Best Local Similarity 83.1%; Pred. No. 3.4e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAGAGTTTGTCTAGTGGTGCAGAAAGCCCAACACTGTTCTTCCAGTGAG 60
|||||
Db 1135 CAATTTAAAATTTCTGTGTAGTGGCTGAAGGTCCTCCACGCTGTTATCTCCCAAGTGAG 1076
|||||
QY 61 TTAGTTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACTTACAGACCCAAAG 118
|||||
Db 1075 TTAAGTTGTACAGAACATCGTCAGCACTAGCACAGTTTACAGAACTTACAGACCCAAAG 1016
|||||
QY 119 GTAC 122
|||
Db 1015 GAAC 1012
|||
RESULT 9
AAS6569/c
ID AAS6569 standard; cDNA; 1318 BP.
XX AC
XX AAS6569;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #1373.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR P-PSDB; ABG01382.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 1373; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
Query Match 64.7%; Score 79.6; DB 23; Length 1318;
Best Local Similarity 83.1%; Pred. No. 3.5e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAGAGTTTGTCTAGTGGTGCAGAAAGCCCAACACTGTTCTTCCAGTGAG 60
|||||
Db 1221 CAATTTAAAATTTCTGTGTAGTGGCTGAAGGTCCTCCACGCTGTTATCTCCCAAGTGAG 1162
|||||
QY 61 TTAGTTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACTTACAGACCCAAAG 118
|||||
Db 1161 TTAAGTTGTACAGAACATCGTCAGCACTAGCACAGTTTACAGAACTTACAGACCCAAAG 1102
|||||
QY 119 GTAC 122
|||
Db 1101 GAAC 1098
|||
RESULT 10
AAF18236/c
ID AAF18236 standard; DNA; 2051 BP.
XX AC
XX AAF18236;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 255.
XX
XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; anti-infective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
XX WO2000055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C.A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
DR P-PSDB; AAB58360.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 1; Page 716-717; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF19424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective, cytostatic, cardioactive,
CC immunomodulatory, muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; anti-infective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
XX

CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.

XX
SQ Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
Query Match 64.7%; Score 79.6; DB 21; Length 2051;
Best Local Similarity 83.1%; Pred. No. 4.1e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAAATTCAGAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTCTTCCAGTGAG 60
Db 1896 CAAATTAAATAATTCGTGTAGTGGTCGAGGTCGCCGCTGTATTCGCCAGTGAG 1837
QY 61 TTAGGTTGTACAGAACGGGTTAGCACTAGCGC--TTGACAGAACCTCAGACCCCAAG 118
Db 1836 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACCTCAGACCCCAAG 1777
QY 119 GTAC 122
Db 1776 GAAC 1773

RESULT 11
ABT17030/c
ID ABT17030 standard; DNA; 28567 BP.
XX
AC ABT17030;
XX
DT 03-APR-2003 (first entry)
XX
DE Human MP21 gene Rac1 SEQ ID No 4.
XX
KW Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
KW cell proliferation disorder; MP21; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003006990-A1.
XX
PD 23-JAN-2003.
XX
PF 10-JUL-2002; 2002WO-US21549.
XX
PR 12-JUL-2001; 2001US-305017P.
PR 10-OCT-2001; 2001US-328491P.
PR 15-FEB-2002; 2002US-357452P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Friedman L, Plowman GD, Belvin M, Li D, Funke RP;
XX
DR WPI; 2003-221779/21.
DR P-PSDB; ABJ19756.
XX

PT Identifying candidate p21 pathway modulator, by contacting an assay
PT system having modifiers of p21 polypeptide or gene with a test agent to
PT provide a reference activity in system and detecting test agent-biased
PT activity -
XX

PS Examples; Page 56-72; 199pp; English.

XX The invention relates to a novel method for identifying a candidate p21
CC pathway modulating agent. The novel method comprises contacting an assay
CC system, comprising a purified MP21 polypeptide (modifier of p21) or
CC nucleic acid, with a test agent under conditions, so that but for the
CC presence of a test agent, the assay system provides a reference activity
CC and detection of test agent-biased activity of the assay system. The
CC novel method of the invention is useful for identifying a candidate p21
CC pathway modulating agent. The invention also includes a method for
CC modulating the p21 pathway of a cell, and a method for diagnosing a
CC disease e.g. cancer in a patient. The identified modulators are useful in

CC diagnosis, therapy and pharmaceutical development. The modulators are
CC useful in a variety of diagnostic and therapeutic applications including
CC angiogenic, apoptotic and cell proliferation disorders. This
CC polynucleotide sequence represents a gene encoding an MP21 protein of the
CC invention.

XX
SQ Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other;
Query Match 64.7%; Score 79.6; DB 25; Length 28567;
Best Local Similarity 83.1%; Pred. No. 1.1e-18;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAAATTCAGAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTCTTCCAGTGAG 60
Db 28012 CAAATTTAAATAATTCGTGTAGTGGTCGAGGTCGCCGCTGTATTCGCCAGTGAG 27953
QY 61 TTAGGTTGTACAGAACGGGTTAGCACTAGCGC--TTGACAGAACCTCAGACCCCAAG 118
Db 27952 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACCTCAGACCCCAAG 27893
QY 119 GTAC 122
Db 27892 GAAC 27889

RESULT 12
ABX46267
ID ABX46267 standard; cDNA; 353 BP.
XX
AC ABX46267;
XX
DT 21-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #11432.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-0960352.
XX
PR 12-JAN-1999; 99US-115707P.
PR 11-JAN-2000; 2000US-0480902.
XX
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WAR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-110599/10.
XX
PT New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle -
XX
PS Claim 2; SEQ ID No 11432; 245pp; English.

XX The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMPD), derived
CC from cattle, and the LMPD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotides
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are; (i) a transformed cell having a nucleic acid
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of

CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMPD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMPD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ Sequence 353 BP; 91 A; 67 C; 85 G; 110 T; 0 other;

Query Match 46.3%; Score 57; DB 25; Length 353;
Best Local Similarity 82.8%; Pred. No. 5.1e-11;
Matches 77; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
QY 1 CAATTGAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTGAG 60
D5 |||||
155 CAATTGAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTG-G 213
QY 61 TTAGTTGTACAGAGCGGTTAGCACTAGCGC 93
D5 |||||
214 GTAAGTTGTACAGAACTTCGTTAGCAGGAC 246

RESULT 13
AAL07296/c
ID AAL07296 standard; DNA; 3740 BP.
XX
AC AAL07296;
XX
DT 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 9984.
DE Human reproductive system related antigen; reproductive system disorder;
XX Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX Homo sapiens.
XX
XX WO200155320-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
PT Disclosure; SEQ ID NO 9984; 1297bp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
SQ Sequence 3740 BP; 1188 A; 752 C; 687 G; 1113 T; 0 other;
Query Match 31.1%; Score 38.2; DB 22; Length 3740;
Best Local Similarity 78.0%; Pred. No. 0.0012;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 8 AAAAGTTTGTCTAGTGTGCAAGGCCCAACACTGTCTTCTGCCAGTGAAGTTAGGT 66
DT 2725 AAAATATTGTCTAGTGTGCAAGGCCCAAGTGTATATCTTCCAGTGTGAAGTT 2667

RESULT 14
ABZ04062/c

ID ABZ04062 standard; DNA; 50 BP.

XX AC ABZ04062;

XX DT 09-JAN-2003 (first entry)

XX DE Human leukocyte gene expression profiling probe SEQ ID NO 4053.

XX KW T7; leukocyte; gene expression profiling; allograft rejection;

XX KW atherosclerosis; congestive heart failure; systemic lupus erythematosus;

XX KW rheumatoid arthritis; osteoarthritis; cytomegalovirus; infection;

XX KW probe; ss.

XX OS Homo sapiens.

XX FN WO200257414-A2.

XX PD 25-JUL-2002.

XX PF 22-OCT-2001; 2001WO-US47856.

XX PR 20-OCT-2000; 2000US-241994P.

XX PR 08-JUN-2001; 2001US-296764P.

XX PA (BIOC-) BIOCARDIA INC.

XX PI Wohlgemuth J, Fry K, Matcuk G, Altman P, Prentice J, Phillips J;

XX PI Ly N, Woodward R, Quertermous T, Johnson F;

XX DR WPI; 2002-636525/68.

XX PT New system for leukocyte expression profiling, diagnosing a disease, or

XX PT monitoring (the rate of) progression of a disease, e.g. atherosclerosis

XX PT or congestive heart failure, comprises diagnostic oligonucleotides

XX PS Claim 1; Page 456; 2038pp; English.

XX CC The invention relates to a system for detecting gene expression, which

XX CC comprises one or two isolated DNA molecules that detect expression of a

XX CC gene, where the gene corresponds to any of 8143 oligonucleotides

XX CC (ABZ00010-ABZ08152) each having 50 base pairs (bp). The system is useful

XX CC for leukocyte expression profiling. It is particularly useful for

XX CC diagnosing a disease, monitoring (rate of) progression of a disease,

XX CC predicting therapeutic outcome, determining prognosis for a patient,

XX CC predicting disease complications in an individual or monitoring response

XX CC to treatment in an individual. The diseases include cardiac allograft

XX CC rejection, kidney allograft rejection, liver allograft rejection,

XX CC atherosclerosis, congestive heart failure, systemic lupus erythematosus,

XX CC rheumatoid arthritis, osteoarthritis or cytomegalovirus infection.

XX SQ Sequence 50 BP; 16 A; 15 C; 10 G; 9 T; 0 other;

Query Match 27.6%; Score 34; DB 24; Length 50;

Best Local Similarity 80.0%; Pred. No. 0.0082;

Matches 40; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 18 TTCTAGTGTGCAAGGCCCAACACTGTCTTCTGCCAGTGAAGTTAGGT 67

DT 50 TTGTAGTGTGCTGAAGGGTCCACGCTGATTCGCCAGTGAAGTT 1

RESULT 15

AAAL6478

ID AAAL6478 standard; DNA; 605 BP.

XX AC AAAL6478;

XX DT 14-JUN-2000 (first entry)

XX

us-09-717-321a-1.rng

Search completed: November 23, 2003, 11:09:27
Job time : 39.0599 secs

XX WPI; 2001-355948/37.
 XX Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels
 XX
 XX Disclosure; Page 7-8; 76pp; English.
 XX
 XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test agent for
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in an reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.
 XX
 XX Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;
 SQ
 Query March 80.3%; Score 1017; DB 22; Length 1017;
 Best Local Similarity 100.0%; Pred. No. 1e-244;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 GAACCTCACAGACCCAAAGGTACCGAGCATGTGTCGCGTGGTGAGGTCTAGAGGG 304
 DB 1017 GAACCTCACAGACCCAAAGGTACCGAGCATGTGTCGCGTGGTGAGGTCTAGAGGG 958
 QY 305 GCGGCATCAATCACATGACAGTGTGGTACTCTGCGAGAGCATGTGTTTCAAGATATC 364
 DB 957 GCGGCATCAATCACATGACAGTGTGGTACTCTGCGAGAGCATGTGTTTCAAGATATC 898
 QY 365 TAAATAGTTTAAATCTGTAAGCCGAGCAGTGTGTTTCAACCCAGTACTAGAAA 424
 DB 897 TAAATAGTTTAAATCTGTAAGCCGAGCAGTGTGTTTCAACCCAGTACTAGAAA 838
 QY 425 ACGAAGGAGCAGTACTGACCTGAGTAAGAGAGGTGAAACAGGACGCACTTCTACT 484
 DB 837 ACGAAGGAGCAGTACTGACCTGAGTAAGAGAGGTGAAACAGGACGCACTTCTACT 778
 QY 485 ATCTACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTGAGACAT 544
 DB 777 ATCTACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTGAGACAT 718
 QY 545 ATTGCTGTTAAGAGGGGTCTTAAAGAAAGCACTTGCTAAGTTAGCAACTGTGAGGA 604
 DB 717 ATTGCTGTTAAGAGGGGTCTTAAAGAAAGCACTTGCTAAGTTAGCAACTGTGAGGA 658
 QY 605 TGCCGATTTTAAATATGAGTCTAACGCCCATCTCGGGAGGAGCAGAGGGGAGGGGG 664
 DB 657 TGCCGATTTTAAATATGAGTCTAACGCCCATCTCGGGAGGAGCAGAGGGGAGGGGG 598
 QY 665 GCTCAAGAGAGACATGATAGATCGGCCATTTGTCTATCTACTGTTTACAGAAATTAAC 724
 DB 597 GCTCAAGAGAGACATGATAGATCGGCCATTTGTCTATCTACTGTTTACAGAAATTAAC 538
 QY 725 CGTTTAAAGAGCTTACCCGTGACACTTTTATTCAGTTGAATTAATCTCATGTAACAATGAG 784
 DB 537 CGTTTAAAGAGCTTACCCGTGACACTTTTATTCAGTTGAATTAATCTCATGTAACAATGAG 478
 QY 785 TGTAAATTAATCTTACTTATATAGTCAAAATAGTGTCTGCTCTCTTTGATGAGCTG 844
 DB 477 TGTAAATTAATCTTACTTATATAGTCAAAATAGTGTCTGCTCTCTTTGATGAGCTG 418
 QY 845 TGTTCACACACTCCACCCAGCACACCCAGCACTAGGAACAGATATCTTCGTTAGAGGCA 904

DB 417 TGTTCACACACTCCACCCAGCACACCCAGCACTAGGAACAGATATCTTCGTTAGAGGCA 358
 QY 905 ACACAGGAGCCAGAGTCTTGTTCAAAGCCTGAGAGCCGCTCAGTGTGTTTATAGAGA 964
 DB 357 ACACAGGAGCCAGAGTCTTGTTCAAAGCCTGAGAGCCGCTCAGTGTGTTTATAGAGA 298
 QY 965 ACTCACTATGAATCAAAAGAGCAGAGCTGTTTACACCCATCTGAGCTACAGTACAAAGTT 1024
 DB 297 ACTCACTATGAATCAAAAGAGCAGAGCTGTTTACACCCATCTGAGCTACAGTACAAAGTT 238
 QY 1025 ACGTAAATGAGCATGGCTGCTAAGTTACAGTGCCTTACATGCGAGCGGTGCTATTAAAGA 1084
 DB 237 ACGTAAATGAGCATGGCTGCTAAGTTACAGTGCCTTACATGCGAGCGGTGCTATTAAAGA 178
 QY 1085 GGCTGTGCTGTGTACACAGGCTCTGGGAGCTACGGGAGGTCTGCCACCTGAGCCAGAA 1144
 DB 177 GGCTGTGCTGTGTACACAGGCTCTGGGAGCTACGGGAGGTCTGCCACCTGAGCCAGAA 118
 QY 1145 GCTGAGTCTTCTTAAAGGACAAAGTCTCTCAACAGCTTAGTGTGCTTACGTTTCTCAGCAC 1204
 DB 117 GCTGAGTCTTCTTAAAGGACAAAGTCTCTCAACAGCTTAGTGTGCTTACGTTTCTCAGCAC 58
 QY 1205 AACGCAACTTAGTTTCAAGAGCTATTTGGCAATTTCTTAATCTGAGCAAGATAGGG 1261
 DB 57 AACGCAACTTAGTTTCAAGAGCTATTTGGCAATTTCTTAATCTGAGCAAGATAGGG 1
 RESULT 3
 AAF18236/c
 ID AAF18236 standard; DNA; 2051 BP.
 XX AC AAF18236;
 XX AC AAF18236;
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polynucleotide sequence SEQ ID 255.
 XX
 KW Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardiocactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotropic; antinefactive; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 XX OS
 XX WO2000055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 DR WPI; 2000-587514/55.
 XX P-PSDB; AAB58360.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 716-717; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardiocactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal

CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAF58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
Query Match 35.4%; Score 447.6; DB 21; Length 2051;
Best Local Similarity 68.2%; Pred. No. 5.4e-102;
Matches 876; Conservative 0; Mismatches 324; Indels 84; Gaps 15;
QY 33 TTTTCTTTTCTTTTATGATTCAAGGATTTTAAAGTACATGCAAAACATCTGC 92
Db TTTTCTTTTCTTTTATGATTCAAGGATTTTAAAGTACATGCAAAACATATGC 1950
QY 93 TAACTGCATAGCAAAAGATCAATGTAAAGCACTCCCAATCTCGCACTGCAATTGA 152
Db TAAATGATAGCAAAAGATCAATGTAAAGCACTCCCAATCTCGCACTGCAATTGA 1890
QY 153 AAAAAGTTTCTTCTAGTGTGCAAGGCCCAACACTGTGTCTTGTCCAGTGTAGGTT 212
Db AAAAATCTTGTGTAGTGTCTGAGGGTCCACGCTGTATCTCGCCAGTGTAGTT 1830
QY 213 GTACAGAACGGCTTAGCTAGCTAGCCG - TTGACAGAACTTCACAGACCCAAAGG - 264
Db GTACAGAACATCTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCAAAGGAATC 1770
QY 265 -----TACCGAGACATGTCGCGGTGGGTGAGGCTAGAGGGGGCGC 309
Db ATAGGCAAGGCACTACAGAGGGCGGTGTCGCGGTGGGGGAGGTAAGA - GGGTCAGT 1711
QY 310 ATCAATCAATGACAGTGTGTGTAATCTGCGCAAGACAGTGTG - 352
Db ATTGGTCAAGTACAGTGTGCGTAATCTGCGCAAGACAGTGTGTTAAGAGGTTTCATAGT 1651
QY 353 TTTCAGATATCTAAATAGTTTAAACCTGTAAAGCGGAGCAGCTGATTTCTACACC 412
Db TTAAGAAATATCTAAATATTTTAAACCTATTAAGCTGCAACACATGATTTTACACCT 1591
QY 413 AGTTACTAGAAACGAGGAGCACTAGTCAGTGTAGTAAAGAGGTCGAAACAGGAA 472
Db AGTTACTAGAAACCTAAGGAAAGCACTTATAGTCTGAATA -----AAGTACATGGAA 1536
QY 473 CGCATTCTCTACTATCTACCAAAAATAATCTCCGAATGATTTATCAGAAAGATCTTATAGTA 532
Db AGCATTCTCTATCTGACAAAATAATCTCTATGATTTATCAGAAAGATTTTATAATA 1476
QY 533 CAGGTGACATATTTGCTGTTAAGAGGGGCTTAAAGAAAGCACTGCTTAAGTAG 592
Db CAAG - GAGGCAATTTGCTCAGTCAGAGAGGGGTTCTATAAGAAAGCACTTACTAAGTAG 1417
QY 593 CAACTGTGAGGATGGCCAGTTTAAATATGACTCAACGCCCATCTCGGGAGGAGACAGCA 652
Db CGACTAACAGAACACAGTTTAAAGATGAATTAATGCCCATTGCGGAGGAGCATGGCA 1357
QY 653 GGGGGAAGGGGGCTCAAGAGAGACACTGATAGATCGGCCATTTGTCATCTACTGTTG 712
Db GGTGTAAGAGAGAGGAAAGCTTAAAGAAACATTTCTCTGATATATACCAACCTTCTTCTCA 1297
QY 713 ACAGAAATTAACCGTTAAAGAGCTTTACCGGTGACACTTTTATTCAGTTGAA ---TTACT 769
Db TCATCTACTGATTTGACAGAAATTAACCTTTTAGAGTTTATACCGGTGACACTTTTCA 1237
QY 770 CCATGTCAATGTAGTGTAAATTAATCTCTACTTTCATA - TTAGTCAAAATACTGTCTGTC 828

Db 1236 CTTTCTCAATCTAGTGA ---AATCTCCACTTGTATTTTGTCAAAATACTGTCTTTG 1181
QY 829 TCCCTTTGATGAGCTGCTTTTACACACTCCACCAGACACCCACGACTAGGAACAGAA 888
Db TCCCTTTGATCA -----CACACACCCACCCGCGCACCCACAGCTA - AACAGAA 1133
QY 889 TACTTCTGTAGAGGCAACACAGGAGCCAGAGTTCTGTTCAAAGCCTGACAGAGCCGCTCA 948
Db TTTCTTCAATTAGAGGAATA -----GCAGTTCTGTTCAAAATCTCGCAAAAGCTGCTCA 1079
QY 949 GCTGTGATTTTATAGAGCACTCACTATGAAATCAAGAGCAGAGCTGTTACACCCATCGTA 1008
Db GAAATCTGCTATGAAATCACAAGAGCTGATCCAAAGAGCTGAGCTGCTACGCTCACTCC 1019
QY 1009 CGTACAGTACAAAGTTTACGTAATGAGCATGGGCTGATAGTTTACAGTGGTTCATATGGC 1068
Db ATTACAGTACAAATGTTATGTCGGGAACACGCTGCTCTAACTCACTGGTGAAGTTCAATGGC 959
QY 1069 AGCGTGTCTAATGAGGAGGCTGTGCTGTACACAGGCTCTGG -----GAGCTACGGGAG 1121
Db AAGCGTTCAATCGGAGGCTGTCTGCTTACGCATCTGAGAACTACATAGGAGAGCAAG 899
QY 1122 GGTCTGCACC - COTGAGCCCCAGAGCTGCAGTCTTCTTAAGGACAAAGTCTCTCAACAGC 1180
Db TGTCTGCACCTCTCTAATCTGAGAGGCTACCGTCTTCTCAAAGACGAAGGTCTTTTGCAAAG 839
QY 1181 TTATGCTTACGTTGTTCTCAGACAAAC ---GCAACTTAGTTCACAGAGTATTTTGGCAAT 1237
Db TTCAAGTGTCTGCTGTTTCTCGGCAACAATGCAAGTGTAGTTCAGAGGTTATTTTGGCAAC 779
QY 1238 TCTTAATCTGAGCAAGAAATAGGGG 1261
Db TCTTAATCTGAGCAAGAAATAGGGG 755
RESULT 4
ABT17030/C
ID ABT17030 standard; DNA; 28567 BP.
XX
AC ABT17030;
XX AC
XX DT
XX 03-APR-2003 (first entry)
DE Human MP21 gene Rac1 SEQ ID No 4.
XX Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
XX cell proliferation disorder; MP21; gene; ds.
XX Homo sapiens.
XX OS
XX WO2003006990-A1.
XX
XX 23-JAN-2003.
XX 10-JUL-2002; 2002WO-US21549.
XX
XX 12-JUL-2001; 2001US-305017P.
XX 10-OCT-2001; 2001US-328491P.
XX 15-FEB-2002; 2002US-357452P.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Li D, Funke RP;
XX WPI; 2003-221779/21.
XX P-PSDB; ABU19756.
XX
XX Identifying candidate p21 pathway modulator, by contacting an assay
XX system having modifiers of p21 polypeptide or gene with a test agent to
XX provide a reference activity in system and detecting test agent-biased
XX activity -
XX

Examples; Page 56-72; 199pp; English.

PS The invention relates to a novel method for identifying a candidate p21
XX pathway modulating agent. The novel method comprises contacting an assay
CC comprising a purified MP21 polypeptide (modifier of p21) or
CC nucleic acid, with a test agent under conditions, so that but for the
CC presence of a test agent, the assay system provides a reference activity
CC and detection of test agent-biased activity of the assay system. The
CC novel method of the invention is useful for identifying a candidate p21
CC pathway modulating agent. The invention also includes a method for
CC modulating the p21 pathway of a cell, and a method for diagnosing a
CC disease e.g. cancer in a patient. The identified modulators are useful in
CC diagnosis, therapy and pharmaceutical development. The modulators are
CC useful in a variety of diagnostic and therapeutic applications including
CC angiogenic, apoptotic and cell proliferation disorders. This
CC polynucleotide sequence represents a gene encoding an MP21 protein of the
XX invention.

SQ Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other;

Query Match 34.9%; Score 441.8; DB 25; Length 28567;
Best Local Similarity 67.4%; Pred. No. 4e-100;
Matches 881; Conservative 0; Mismatches 342; Indels 84; Gaps 15;

QY 10 TTTTTCAGGTTCCAAAGACATTTTTCATGATTCAGGATTTATTA 69
DB 28148 TTGTTCAATCAAGTCTCAAAACACCGAGTATTCAGGATTTATTA 28089
QY 70 GTCATACATGCAAAACATCTGTAATGCAATAGCAAAAGATCAATGTAACACATCC 129
DB 28088 GTCATACATGCAAAACATCTGTAATGCAATAGCAAAAGATCAATGTAACACATCC 28029
QY 130 ACATTCGCAACGTCAATTTGAAAAGTTGTTCTAGTGGTGAAGGCCCAACATG 189
DB 28028 ACAATTCGCAACGTCAATTTGAAAAGTTGTTCTAGTGGTGAAGGCCCAACATG 27969
QY 190 TGTTCCTGCCAGTCAGTGTAGTCTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAA 247
DB 27968 TATTCCTGCCAGTCAGTGTAGTCTACAGAACCATCTGTCAGCACTAGCACTTTACAGAA 27909
QY 248 CCTCACAGACCCAAAGG-----TACCGGAAGCATGTGTCGCGGT 286
DB 27908 CCTCACAGACCCAAAGGAAACATCAATAGGCAAGCGACTACAGAGCGGTGTGTCGCGGT 27849
QY 287 GGTGAGGCTAGAGGGGGGGGATCAATCATGACAGTGTGCTGCTGCTGCAAGACA 346
DB 27848 GGGCGAGGTTAAAGA-GGGTCAGTATGCTCAAGTCAGAGTGTGCTGCTGCAAGACA 27790
QY 347 GTGATGTTTCAGAA-----TATCTAAATAGTTTAAACCTGTAAGC 389
DB 27789 GTGATGTTAAGAGTTTCATAGTTTAAAGATTTCTAAATATTTTAAACCTGTAAGC 27730
QY 390 CGCAGCAGCTGATTTCTACACCCAGTTACTAGAAAAGCAAGGAGCACTAGTCAGCTGA 449
DB 27729 TGCAACACATGATTTTACACCTAGTTACTAGAAAACCTAAGGAAGCACTTATAGCTCT 27670
QY 450 GTAAGGAAGGTGAAACAGAGAACGCACTTCTACTATCTACCAAAAATCTCCGAATGC 509
DB 27669 GAATA-----AAGTAAACATGGAAGCACTTTTACTAATCGCAAAAAAACCTTCAATGC 27615
QY 510 ATTATCAGAAGATCTTTATAGTCAGGTCAGACATATTTGCTGTTAAGAGGGGGTCTTA 569
DB 27614 ATATCAGAAGATTTTATATATACAG-GAGGCAATATGCTCAGTCAGAAGGGGTCTAT 27556
QY 570 AAGAAAAGCACTTCTAAGTTAGCAACTGTGAGGATGCCAGTTTAAATATGGAATCAAC 629
DB 27555 AAGAAAAGCACTTCTAAGTTAGGCACTTAACAGAACCAACAGTTTAAAGATGAATTAAT 27496
QY 630 GCCCCATCTGGGAGGACAGCAGGGGAGGGGGCTCAAGAGAGACACTGATAAGATC 689
DB 27495 GCCCAATTTGGGAGGATGAGGCTGTAAGAGAAAGGAAAGCTTAAGAAAACATTTCC 27436
QY 690 GGCCATTTGTCTACTGTTTGCAGAAATTAACCGTTTAAAGCTTTACCCGTGACAC 749

DB 27435 TGATAATACCAACCTTTCTTTCATCATCTACTGCAATTTGCAGAAATTAACCTTTTAGAG 27376
QY 750 TTTTATTCAGTTGAA---TTACTCCATGTACAATGTAGTGTAAATTAATCTCTACTTTCAT 806
DB 27375 TTTTACCGGTGACACTTTCTTCTGTGTAATGTAGTGTGTA---AATCTCCATCTCGT 27320
QY 807 A-TTAGTCAAAATACTGTCTGTCTCTCTTTGTAGTGTGCTGTTTTCACACATCTCCACCCAG 865
DB 27319 ATTTGTCAAAATACTGTCTGTCTCTTTGCTTTGATCA-----CACACACCCACCCGG 27270
QY 866 CACACCCACGACTAGGACAAATACCTCTGTAGAGGCAACACAGGAGCCAGATTTCTGT 925
DB 27269 CACACCCACAGCTA--AACAGAAATTTCTTATTAGAGGAAATA-----GCAGTTCTGTTT 27218
QY 926 TCAAGAGCTGCAAGACCGGTGCTGTGTTATTTTAGAGAACTCACTATGAATCAAGAG 985
DB 27217 AAAATCTCCGCAAAAGCTGTGTCAGAAACTCGCTATGAATCAAAAGACTGATCCAAAG 27158
QY 986 CAGAGCTGTTTACACCCATCTGTGAGCTACAGTACAAAGTTAGCTAATGAGCATGGGCTGAT 1045
DB 27157 AGCTGAGCTGTCTACGCTCACTCCATTTACAGTACAAATGTTATGTCGGGAACACGTGCTGT 27098
QY 1046 AAGTTACAGTGCCTTACATGCGCAGCGTGTCAITTAAGAGGCTGTGCTGTGTACACAGT 1105
DB 27097 AACTCACTGTGTGAGTTCAATGGCAACGCTTCATTCGGGAGGCTGTTCTGTCTTACGCATC 27038
QY 1106 CTGG-----GAGTACGGGAGGCTGTGACC--CTGAGCCCAAGAGCTGAGTCTTCT 1157
DB 27037 TGAGAACTACATAGAGAGCAAGTGTCTGACCTCTTAACCTGCAAGACTACCGTCTTCT 26978
QY 1158 TAAGGCAAAAGTCTCTCAACAGCTTGTAGTGTGTCTGCTGTCTGCTGTCTGCAACAAC---GCAACTT 1214
DB 26977 CAAGAGCAAGGCTCTTTCGCAAGTTCAGTGTCTGCTGTCTGCTGCAACAACAATGCAAGT 26918
QY 1215 AGTTCAAGAGTATTTTGGCAATCTTAATCTGAGCAAGATAGGGG 1261
DB 26917 AGTTCAGAGGATTTTGGCAACTCTTAATCTGAGCAAGATAGGGG 26871

RESULT 5

ABK83623/c
ID ABK83623 standard; cDNA; 1232 BP.

XX AC ABK83623;

XX XX 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #194.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

XX 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX

DR WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression

PT of genes associated with granulocyte activation, which serves as

PT diagnostic markers that is useful for monitoring disease states and

PT drug toxicity

XX

PS Claim 1; SEQ ID No 194; 114pp; English.

XX

CC The invention relates to detecting (M1) granulocyte (GC) activation

CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by

CC DNA chip analysis as given in the specification, and comparing

CC the expression level to an expression level in an unactivated

CC GC, where differential expression of Gs is indicative of GCA.

CC Also included are modulating (M2) Gs by contacting GC with an agent

CC that alters the expression of at least one gene in Gs; (2) screening (M3)

CC for an agent capable of modulating GCA or an inflammation (especially

CC chronic) in a tissue, an allergic response in a subject, exposure of a

CC subject to a pathogen or sterile inflammatory disease using the

CC gene expression profile; (3) detecting (M4) an inflammation (especially

CC chronic) in a tissue, an allergic response in a subject, exposure of a

CC subject to a pathogen or sterile inflammatory disease, by detecting the

CC level of expression in a sample of the tissue of gene(s) from Gs, where

CC the level of expression of the gene is indicative of inflammation;

CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,

CC an allergic response in a subject, exposure of a subject to a pathogen

CC or sterile inflammatory disease, by contacting a tissue having

CC inflammation with an agent that modulates the expression of gene(s)

CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for

CC modulating Gs; M3 is useful for screening an agent capable of modulating

CC GCA preferably in an inflammation in a tissue; M4 is useful for

CC detecting an inflammation (especially chronic) in a tissue, an allergic

CC response in a subject, exposure of a subject to a pathogen or sterile

CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,

CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal

CC reperfusion injury, ARDS, adult respiratory distress syndrome,

CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,

CC periodontal disease; also bacterial infection, viral infection,

CC parasitic infection, protozoal infection, fungal infection and M5 is

CC useful for treating one of the above conditions. The present

CC sequence represents a gene differentially expressed in granulocytes.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;

Query Match 33.2%; Score 420.2; DB 24; Length 1232;

Best Local Similarity 67.5%; Pred. No. 3.3e-95;

Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;

QY 53 ATTCAGAGATTATTAAGTCATACATGCAAAACATCTGCTAATCTAGCAAAAGAT 112

DB 1228 WTCCAGAGATTATTAAGTCATACATGCAAAACATCTGCTAATCTAGCAAAAGAT 1169

QY 113 CAATGTAATAAAGCACTCCCAATCTCAACTGCAATGTAATAAAGTTCTTCTAGTGGT 172

DB 1168 CAATGTAATAAAGCACTCCCAATCTCAACTGCAATGTAATAAAGTTCTTCTAGTGGC 1109

QY 173 CGAAGGCCCAACACTGTGTTCTTTCGCAAGTGTAGTGTGTACAGAACGGCGTTAGCAC 232

DB 1108 TGAAGGGTCCACGGTGTATCTTCGCGAGTGTAGTGTGTACAGAACATCGTCAGCAC 1049

QY 233 TAGCGC--TTGACAGAACCTCAGAGACCCAAAG-----TACCG 269

DB 1048 TAGCACAGTTTACAGAACCTTCACAGACCCCAAGAGAACATCAATAGGCAAGCACTACAG 989

QY 270 GAAGCATGTCTCCGCGTGGGTGAGTCTACAGAGGGGGGGCGGCATCAATCACATGACAGTGT 329

DB 988 GAGCGTGTCTCCGCGTGGCGAGGTAAGA-GGGTCAGTATTGGTCAAGTCAGACAGTGT 930

QY 330 GGTAAGTCTGGCAAGACAGTGTATG-----TTTCAGAAATATCAAAATAG 372

DB 929 GGTAAATCTGCAAGCAGTGATGTTTAAAGAGGTTCATAGTTTAAAGAAATTATCTTAAATAT 870

QY 373 TTTTAAATCTGTAAGCCCGCAGACGTGATTTCTACACCCAGTTTACTAGAAAACGAGGG 432

DB 869 TTTTAAATCTATAAAGCTGCAACACATGATTTTACACCTAGTTTACTAGAAAACCTAAGGA 810

QY 433 AAGCACTAGTCAAGTGTAGTAAGAGAGGTGAAACAGAGAACGCACTTCTACTATCTACCA 492

DB 809 AAGCACTTATAGTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAATCGACA 755

QY 493 AAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTGACACATATCTCTCG 552

DB 754 AAAAAACCTTCTAATGCAATTATCAGAAAGATTTATAATACAAG-GAGGCATATTGCTCA 696

QY 553 TTAAGAGGGGGTCTTAAGAAAAGCACTGTAAAGTTAGCACTGTGAGGATGGCCAGT 612

DB 695 GTCAGAGGGGGTCTTATAAGAAAAGCACTTACTAAGTTAGCGACTAACAGAACCAACAGT 636

QY 613 TTAATATGGACTCAACGCCCATCTGGGAGGAGCAGCAGAGGGGGAAGGGGGCTCAAGA 672

DB 635 TTAAGATGAATTAATGCCCATTGTTGGGAGGCATGGCAGGTGTAAAGAAAGAAAG 576

QY 673 GAGACACTGATAAGATCGGCCCATTTGTCTACTTCTGTTTACAGAAATTAACCGTTAAAA 732

DB 575 CTTAAGAAAACATTTCTCTGATAATACCAACCTTCTTCTCATCATCTACTGCAATTCACAG 516

QY 733 AGCTTTACCGGTGACACTTTTATTCAAGTTGAA---TTACTCCATGTACATAGTGTAA 789

DB 515 AAATTAACCTTTTAGAGTTTTTACCCGTGACACTTTCATTCTTTGACATAGTAGTGA- 457

QY 790 ATTAACTCTACTTTCATA-TTAGTCAAAATACTGCTGCTCTCTCTTGTGACGCTCGTGT 848

DB 456 ---AATCTCCACTTCTGATTTTGTCAAAATACTGCTTCTTGTCTTTGATCA----- 409

QY 849 TCACACACTCCACCCAGCAGACCCAGCATAGAGAACAGAAATCTTCTGTAGAGGCAACAC 908

DB 408 -CACACACCCACCCCGGCACACCCACAGCTA--AACAGAAATCTTCTATTAGAGGAAATAG 352

QY 909 AGGAGCCAGAGTCTGTTTCAAGCCCTGCAAGAGCCGTCAGCTGTTATTTAGAGAACTC 968

DB 351 -----CAGTTCTGTTTCAAAATCTCCGAAAGCTGCTGACAGAAACTCGCTATGAATCA 298

QY 969 ACTATGAATCAAAAGAGCAGAGCTGTTACACCCATCTGACGTACAGTACAGAACTGCT 1028

DB 297 CAAGACTGATCCAAAGAGCTGAGCTGCTAGCTCACTCCATTACAGTACAATGTTATGT 238

QY 1029 AATGACATGGGCTGATAGTTACAGGTGCTTACAGGTGCTTACAGGTGCTTACATTAAGGAGGCT 1088

DB 237 CGGGAACACGTGCTGCTAACTCACTGCTGAGTTCATGTCGCAACGCTTCAATTCGGAGGCT 178

QY 1089 GTGCTGTGTACACAGGTCTGG-----GAGCTACGGGAGGCTCCACC-CCTGAGCCC 1140

DB 177 GTTCTGCTTTACGCATCTCAGAACTACATAGGAGAGCAAGTGTCTGACCTCTCTAATCTG 118

QY 1141 AGAAGCTGCAAGTCTTCTTAAGGACAAAGTCTCTCAACAGCTTGTAGTGTACGTGTTCTCA 1200

DB 117 AGAAGCTACCGTCTTCTCAAGAGAGGAGTCTTTGCTAAGTTCAGTCTCGGTGTTCTCG 58

QY 1201 GCACAC--GCACTTAGTTCAGAGTATTTTGGCAATTTCTTAATCTGAGCAAGA 1254

DB 57 GCACAACTAGCAGTGTAGTTTCAAGAGGATTTTGGCACTCTTAATCTGAAACAAGA 1

RESULT 6

ABN95134/c

ID ABN95134 standard; DNA; 1232 BP.

XX

XX ABN95134;

XX AC

DT 13-AUG-2002 (first entry)

XX DE Gene #1632 used to diagnose liver cancer.

```
XX Gene: liver cancer; ds: hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytosolic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX Homo sapiens.
OS WO200229103-A2.
PN 11-APR-2002.
PD 02-OCT-2001; 2001WO-US30589.
PP 02-OCT-2000; 2000US-237054P.
PR (GENE-) GENE LOGIC INC.
PA
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX Diagnosing and detecting the progression of liver cancer,
XX hepatocellular carcinoma or metastatic liver tumor in a patient,
XX involves detecting the level of expression of two or more genes in a
XX liver tissue sample
XX Claim 1; SEQ ID NO 1632; 298pp; English.
XX
XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumour in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytosolic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
XX
XX Query Match 33.2%; Score 420.2; DB 24; Length 1232;
XX Best Local Similarity 67.5%; Pred. No. 3.3e-95;
XX Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;
XX
XX QY 53 ATTCAGGATTTATTAGTCATACATCGAAACATCTGCTAAGTCTAGTAAAGAT 112
XX Db 1228 WTCAGGATTTATTAGTCTATCATCGAAACATCTGCTAAGTCTAGTAAAGAT 1169
XX
XX QY 113 CAATGTAAAAAAGACTCCACAACTCTGCAATGTCTGCAATGTGAAAGATTTCTCTAGTGGT 172
XX Db 1168 CATGTAAAAAGACTCCACAACTCTGCAATGTCTGCAATGTGAAAGATTTCTCTAGTGGT 1109
XX
XX QY 173 CGAAGGCCCAACACTGTGTCTTTCGAGTGTAGTGTGTGACAGAACGGCGTTAGCAC 232
XX Db 1108 TGAAGGTCGCCAGCTGTATTCTCGCAGTGTAGTGTGTGACAGAACGGCGTTAGCAC 1049
XX
XX QY 233 TAGGCG--TTGACAGAACCTCAGAGCCCAAGG-----TACCG 269
XX Db 1048 TAGCAGAGTTTACAGAACCTCAGAGCCCAAGGACATCAATAGGCAAGCGACTACAG 989
XX
XX QY 270 GAAGCATGTCTCGCGTGGTGTAGGTTAGAGGGGGCGGCATCAATCAGATGACAGTGT 329
XX Db 988 GAGCGGTGTCTCGCGTGGCGAGGTAAGA-GGGTCAGTATTGGTCAAGTGACAGTGC 930
XX
XX QY 330 GGTAATCTGCAAGACAGTGTATG-----TTTCAGAAATATCTAAATATAG 372
XX Db 929 GGTAATCTGCAAGACAGTGTATGAGAGGTTTCATAGTTTAAAGATTATCTAAATAT 870
XX
```

```
QY 373 TTTAAAAAAGCTGTAAAGCCGACGACGTGATTTCTTACACCCAGTACTAGTAAACGAAGG 432
Db 869 TTTAAAAAAGCTGTAAAGCCGACGACGTGATTTCTTACACCCAGTACTAGTAAACGAAGG 810
QY 433 AAGCACTAGTGTAGTGTAAAGGAGGTGAAACAGGAAACGCACTTCTTCTATCTATCTACCA 492
Db 809 AAGCACTTATTAGTCTCTGAATA-----AAGTAAACATGGAAGCACTTTTACTTAATCGACA 755
QY 493 AAAAATCTCCGAATGCAATTATCAGAAAGATCTTATAGTACAGGTACAGATATTGCTG 552
Db 754 AAAAATCTCTTATGCAATTAACAGAAATTTTATAATACAG-GAGCAATTTGCTCA 696
QY 553 TTAAGAAGGGGTCTTAAGAAGAAAGCACTTGTAGTTAGCAACTGTGAGGATGCGCAGT 612
Db 695 GTCAGAAGGGGTCTTATAGAAGAAAGCACTTACTAAGTTAGCGACTTAAACAGAACACCACT 636
QY 613 TTAATATGGACTCAACGCCCACTCTGGGAGGAGACAGCAGGGGGAAGGGGGCTCAAGA 672
Db 635 TTAAGATGAATTAAATGCCCAATTTGGGAGGAGCATGCGAGGTGTAGAGAAAGGAAAG 576
QY 673 GAGACACTGATAAGATCGGCCATTTGTCTACTCTGTTTGTAGCAGAAATTAACCGTTAA 732
Db 575 CTTAAGAAACATTTCTCTGATTAACCAACCTTTCTTTCATCATCTACTGCAATTTGACAG 516
QY 733 AGCTTTACCCGTGACACTTTTATTTCAGTTGAA---TTACTCCATGTACAAATGTAGTAA 789
Db 515 AAATTAACCTTTTAGAGTTTTTACCCTGTGACACTTCACTTCTTGTACAATGTAGTGA- 457
QY 790 ATTAATCTCTACTTCAATA-TTAGTCAAAAATCTGTCTGTCTCTTGTAGTACGCTGTT 848
Db 456 ---AATCTCCACTCTGATTTTGTCAAAATATACTCTTCTTGTCTCTTGTATCA- 409
QY 849 TCACACTCTCCACCGACACACCCACGACTAGGAAAGAAATCTCTGTTAGAGGCAACAC 908
Db 408 -CACACACCCACCCCGGACACCCACAGCTA--AACAGAAATCTCTTATAGAGAAATAG 352
QY 909 AGGAGCCAGAGTTCTGTTCAAGCCTCGAAGCCGCTGAGTCTGCTGTTATTTTAGAGAACTC 968
Db 351 -----CAGTTCTGTTCAAAATCTCCGAAAGCTGTCAGAAACTCGCTATGAATCA 298
QY 969 ACTATGAATCAAGAGACAGAGCTGTTACACCCATCTGAGTACAGTACAGTACAAAGTTAGCT 1028
Db 297 CAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTCCATTTACAGTACAAATCTTATGT 238
QY 1029 AATGAGCATGGGTGATAGTTTACAGTGTGTTTACATGCGAGCGTGTCAATTAAGGAGCT 1088
Db 237 CGGAAACAGCTGCTGCTGCTTACTCTGCTGAGTTCATGGAACCGCTTCACTCGGAGGCT 178
QY 1089 GTGCTGTGTCAACGGTCTGG-----GAGCTACGGGAGGCTGTGACC-CCTGAGGCC 1140
Db 177 GTTCTGCTTTTACGATCTGAGAACTACATAGGAGAGCAAGTGTGCTGACCTCTCTAATCTG 118
QY 1141 AAGAGCTGCACTCTTCTTAAAGCAAAAGTCTCTCAACAGCTTGTAGTCTTACGTTCTCA 1200
Db 117 AGAAGTACCGCTCTTCTCAAGAGCAAGGTCCTTTGCTAAGTTCAGTCTGCTCGGTCTCTCG 58
QY 1201 GCACAA--GCAACTTAGTTTCAAGGATTTTGGCAATTTCTTAATCTCTAGCAAGA 1254
Db 57 GCACAAATGCACTGTAGTTTCAAGAGGATTTTGGCAACTCTTAATCTCTGAACAAGA 1
XX
```

RESULT 7

AAH22399/c

ID AAH22399 standard; DNA; 348 BP.

XX AAH22399;

AC AAH22399;

XX 22-AUG-2001 (first entry)

XX Human rac1 gene related nucleotide sequence #3.

XX Identification; toxic; hepatotoxic; differential gene expression;

KW

Db 1254 CAATGTAARAACACTCCACRAATCTGCACATCTGCAATTTAAATAAATCTGTTGTAGTGGC 1195
Qy 173 CGAAGGCCCAACACTGTTGCTTCTGCGAGTGTAGTTGTTAAGAACGGCGTTAGCAC 232
Db 1194 TGAAGGGTCCACGCTGTTATCTGCGCAGTGTAGTTAGTTGTCAGAACATCGTCAGCAC 1135
Qy 233 TAGCGC--TTGACAGAACCTCACAGACCCCAAGG-----TACCG 269
Db 1134 TAGCACAGTTTACAGAACCTCACAGACCCCAAGGAAACATCAATAGGCAAGCGACTACAG 1075
Qy 270 GAACATGTGTCCGCGTGGGTGAGGTCTAGAGGGGGGGCATCAATCACATGACAGTGT 329
Db 1074 GAGCGGTGTGTCGCGTGGCGAGGTAAAGA-GGGTCAGTATTGTTCAAGTGACAGTGC 1016
Qy 330 GGTACTCTGGCAGACAGTGAATGTTTC-----GAATATCTAAATAG 372
Db 1015 GGTAAATCTGGCAGACAGTGTATGTTAAGAGGTTTCATAGTTTAAAGAAATATCTAAATAT 956
Qy 373 TTTAAAAACTGTAAAGCGGACGACGTGATTTCTACACCCAGTTACTAGRAAAGCAAGGG 432
Db 955 TTTAAAACTATAAAGCTGCACACATGATTTTACACCTAGTTACTAGAAACTAAGGA 896
Qy 433 AAGCACTAGTCAGTGTAGTAAAGGAAGGTGAAACAGSAGCAGCTTCTACTATCTACCA 492
Db 895 AAGCACTTATTAGCTCTGAATAA-----GTAACATGGAAGCACTTTTACTATTCGACA 841
Qy 493 AAAAAATCTCGAATGATTAATACAGAAGATCTTATAGTACAGTGCAGACATATGCTCG 552
Db 840 AAAAAACCTTCTAATGCAATATCAGAAGATTTTATAACAAG-GAGGCATATTGCTCA 782
Qy 553 TTAAGAGGGGGTCTTAAGAAA-----AGCACTTGTAGTTAGCAACTCTGAGGATGG 607
Db 781 GTCAAGAGGGGTCTTTTAGGAAAGGCACCTTTACTAAGTTAAGCCACTAACAGAACAA 722
Qy 608 CCAAGTTTAAATATAGACTCAA-----CGCCCACTCTGGGAGGAGCAGCAGGGGGAAGGGG 664
Db 721 CCAAGTTTAAAGCATGAATTAATGSCCCAAATTTGGGGGAGGCAATGGCAGGGTTAAGAGA 662
Qy 665 GCTCAAGAGA-----GACACTGATAAGATCGGCCANTTTGTCAT 702
Db 661 AAGGAAAAGGCTTAAGAAAAATTTCTGTGATAATACCAAACTTTCTTTCATTCATCTA 602
Qy 703 CTACTGTTTACAGAAATTAACCGTTTAAAGCTTTTACCCTGTCAGACTTTTATTCAGTTG 762
Db 601 CTGCAITGGACAGAAATTAACCGTTTATAGAGTTTTTACCCCGTGACCACTTTTCATTC 542
Qy 763 AATTACTCCATGTACAAATGATGTAAATTAATCTCTACTTTCATATTAGTCAAAATACTG 822
Db 541 CTGTGTACCAATGGTAGTGGTAATAATCTCCCACTCCGTATTTTGTCCA--AAATTACTG 485
Qy 823 TCTGTCTCCTTTGATGACGTGGTGTTCACACACTCCACCAGCAGACCCACGACTAGGA 882
Db 484 TCTTTGTCCTTTGA-----TCCACCACACACACCCCGGACACCCCGACACGACTAAA 430
Qy 883 ACAGAATACCTCGTTAGAGGCAACACAGGAGCCAGAGTCTGTTTCAAAGCCTGCAGAGC 942
Db 429 CCGAATCTCTTATAGAGAAATACCCAGTTCTGTTCAAATCTCCCGCAAAAGCTGG 370
Qy 943 CGGTACAGTGGTATTTTAGAGAACTACTATGAAATCAAGAGCAGAGCTGTTTACACCCA 1002
Db 369 TCAGAAAACCTCGCTATGAATCAAAAGACTGATCCAAAGAGCTGAGGCTGCTACGCTCA 310
Qy 1003 TCGTGACGTACAGTACAAAGTTAGTAATGAGCATGGGCTG-ATAAGTTACAGGTGGCT 1061
Db 309 --CTCCATTACAGTACAAATGTTATGTGCGGAACACGTCGCTGCTAACTCACTGCTGAT 252
Qy 1062 ACATGGCAGGTGTCAATTAAGAGGCTGTGCTGTCTACACGGTCTCGGA-----GC 1113
Db 251 CAATGGCAACGTTTCAITCGGGAGGCTGTTCTGCTTTACGCATCTGAGAACTACATAGGA 192
Qy 1114 TAGGGAGGGTCTGCACC-CCTGAGCCCAAGAAGTGCAGTCTTCTTAA--GGACAAAAGTC 1170
Db 191 GAGCAAGTGTCTGCACCTCCTAACTGCAGAGGTTACCGTCTTCTCAATGACGAGGTC 132

Qy 1171 TCTCAACAGCTTAGTGT--TACGTGTTCTCAGCAACGCAACTTAGTTCAAGGTAT 1228
Db 131 TTTGCAAGTTTCAGTGTGCGTGTTCGCGCACCAATGCAGTGTAGTTCAAGGTAT 72
Qy 1229 TTTGGCAATCTTAACTGTGACAGAAATAGGG 1261
Db 71 TTTGGCACTCTTAACTGTGACAGAAATGGGG 39

RESULT 9

AAH22398/c

ID AAH22398 standard; DNA; 1022 BP.

XX AC AAH22398;

XX DT 22-AUG-2001 (first entry)

XX DE Human rac1 gene related nucleotide sequence #2.

XX KW Identification; toxic; hepatotoxic; differential gene expression;
NSAID; non-steroidal antiinflammatory drug; ds.

XX OS Homo sapiens.

XX EN WO200138579-A2.

XX PD 31-MAY-2001.

XX PF 21-NOV-2000; 2000WO-US32049.

XX PR 22-NOV-1999; 99US-0166923.

XX PR 18-FEB-2000; 2000US-0183531.

XX PR 20-NOV-2000; 2000US-0717321.

XX PA (CURA-) CURAGEN CORP.

XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX DR WPI; 2001-355948/37.

XX PT Screening hepatotoxic agent comprises contacting test cell population
expressing RISKMARKER or INJURYMARKER with agent, comparing expression
with reference population and identifying difference in expression
levels

XX PS Disclosure; Page 7-8; 76pp; English.

XX CC The present invention describes a method of screening a test agent for
hepatotoxicity. The method comprises: (a) providing a test cell
population comprising a cell capable of expressing one or more nucleic
acid sequences selected from the group consisting of RISKMARKER 1-8
and INJURYMARKER 1-10; (b) contacting the test cell population with a
test agent; (c) measuring expression of one or more of the nucleic
acid sequences in the test cell population; (d) comparing the
expression of the nucleic acid sequence in the test cell population to
the expression of the nucleic acid sequence in an reference cell
population comprising at least one cell whose exposure status to a
hepatotoxic agent is known; and (e) identifying a difference in
expression levels of the RISKMARKER or INJURYMARKER sequences, if
present, in the test cell population and reference cell population.
The method is useful for identifying a hepatotoxic agent. The present
sequence is given in the exemplification of the present invention.

XX SQ Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;

Query Match

Best Local Similarity 23.0%; Score 290.6; DB 22; Length 1022;

Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

Qy 265 TACCGAAGCATGTCCGGTGGTGGTGTAGAGGGCGGCATCATCATGACA 324

Db 1000 TACAGGAGCGGTGTCCCGTGGGAGGTAAGA-GGGTCAGTATTGTCAGGTGACA 942

Db 388 TGCACCTGCAATTGAAAAAGTTTGTCTTAGTGGTTGAAAGGCCCAACACCTGCATCTT 329
 QY 196 TGCCAGTCAGTAGTGTGTACAGACGGGTTAGCACTAGCCCTTGACAGAACCTTCACAG 255
 Db 328 TGCCAGTCAGTAGTGTGTACAGACGGGTTAGCACTAGCCCTTGACAGAACCTTCACAG 269
 QY 256 ACCCAAAAGG-----TACCGGAAGCATGTGTCCCGGTGGGTGAGG 294
 Db 268 ACCCAAGGAACATCTTAGGCAGACCATGACAGGAAGGCTGTCTCCAGTGGCGGAGG 209
 QY 295 TCTAGAGGGCGGCATCAATCACATGACAGTGTGTGTCTGTGCAAGACAGTGAATGT 354
 Db 208 TCTAGA---GCAGCATTAGTACATGACAGTGTGTGTCTGTGCAAGACAGTGAATGT 152
 QY 355 TCAGATATCTAAATAGTTTAAACCTGTAAGCCGACAGCTGATTTCTACACCCAG 414
 Db 151 TAAGATATCTAAATAGTTTAAACCCGTAAGCTGCAACACATGATTTGACACCTAG 92
 QY 415 TTACTAGAAACGAAGGGAAGCACTAGT-CAGCTGAGTAAAGGAGGTGAAACAGGAAC 473
 Db 91 TTGCTAGGAACCTAAGGAAGCACTAAGTCTGTGAGTAAAGTAAAGGCAACAGGACG 32
 QY 474 GCATCTTCTACTATCTACCAAAAAA 498
 Db 31 CACTTCTACTGAGCTACCAAAAAA 7

RESULT 11

AA89694/c

ID AA89694 standard; cDNA; 422 BP.

XX AC AA89694;

XX XX

XX 08-JAN-2001 (first entry)

XX Mouse Rab2 nucleotide sequence #1.

XX XX

XX Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
 KW allergic; antiallergic; antiallergic; antiallergic; antiallergic;
 KW vulnary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
 KW digestion disorder; wound healing disorder; gene therapy; ss.
 XX Mus sp.

XX OS

XX WO200043419-A2.

XX XX

XX 27-JUL-2000.

XX XX

XX 20-JAN-2000; 2000WO-US01431.

XX PF

XX 20-JAN-1999; 99US-0116534.

XX PR

XX 26-JAN-1999; 99US-0117274.

XX PR

XX 26-JAN-1999; 99US-0117308.

XX PR

XX 26-JAN-1999; 99US-0117309.

XX PR

XX 01-FEB-1999; 99US-0118172.

XX PR

XX 01-FEB-1999; 99US-0118177.

XX PR

XX 01-FEB-1999; 99US-0118179.

XX PR

XX 09-FEB-1999; 99US-0119286.

XX PR

XX 11-FEB-1999; 99US-0119998.

XX PR

XX 11-FEB-1999; 99US-0119759.

XX XX

XX (RIGE-) RIGEL PHARM INC.

XX PA

XX LUO Y;

XX XX

XX WPI; 2000-482908/42.

XX XX

XX New nucleic acids encoding Exo proteins which are useful in the
 PT diagnosis, treatment or prevention of exocytosis-mediated disorders
 PT such as asthma, inflammation and allergies -
 XX

PS Disclosure; Page 271-272; 305pp; English.

XX The present sequence encodes a polypeptide which is associated with
 CC the exocytosis pathway. cDNA molecules encoding proteins involved in
 CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
 CC screening. Novel proteins, termed Exo proteins, have been identified that
 CC interact with known exocytosis-associated proteins such as GS27, alpha
 CC snap, unc18-1, vamps, snap-23, and the rab family of proteins.
 CC Exo proteins and their agonists and antagonists are useful in the
 CC diagnosis, treatment or prevention of exocytosis-mediated disorders
 CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
 CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC diabetes, digestion disorders and wound healing disorders.
 CC The nucleic acids, antagonists or agonists of Exo proteins are useful
 CC in gene therapy. The nucleic acids are also useful for generating
 CC transgenic or knock-out animals which can be used in the
 CC development and screening of therapeutically useful reagents.

XX Sequence 422 BP; 95 A; 92 C; 83 G; 148 T; 4 other;

Query Match 17.0%; Score 215.8; DB 21; Length 422;

Best Local Similarity 77.0%; Pred. No. 3.5e-44;

Matches 322; Conservative 0; Mismatches 70; Indels 26; Gaps 4;

QY 104 GCAAAAGATCATGTAAACACT-CCACAATTCGCAACTGTCAATTGAAAAAGTTTG 162
 Db 421 GCAAAAGATCAATGTAAACACTCCCACTTCTGCAACTGTCAATTGAAAAAGTTTG 362
 QY 163 TTCTAGTGGTGAAGGCCCAACTGTGTTCTTCCAGTCAGTGTGTTGACAGACG 222
 Db 361 TTCTAGTGGTGAAGGCCCAACTGTGTTCTTCCAGTCAGTGTGTTGACAGACG 302
 QY 223 GCGTTAGCACTAGCGCTTGACAGAACCTCAGACACCCCAAGG----- 264
 Db 301 GCGTTAGCACTAGCGCTTGACAGAACCTCAGACACCCCAAGGACATCTNTAGCGACG 242
 QY 265 ---TACCGGAAGCATGTGTCGGGTGGTGGTGTAGAGGGGGCGGCATCAATCACATG 321
 Db 241 CATGACAGGAAGCGTCTGTCCACGTCGGCGAGGTTAGA---GGCAGCATTAGTCACATG 185
 QY 322 ACAGTGTGTTACTCTGCGAAGCAGTGATGTTTCAGAAATATCTAAATAGTTTAAAAAC 381
 Db 184 ACAGTGTGTTACTGTTTAGAGGTTTCATAGTTTAAAGTATCTTAAATAGTTTAAAAAC 125
 QY 382 TGTAAAGCCGACGACGTCGATTTCTACACCCAGTTACTAGAAAACGAAGGAAGCACTAG 441
 Db 124 CGTAAAGCTGCAACACATGATTTGCACACCTAGTTGCTAGAAACTAAGGAGCACTAAC 65
 QY 442 T-CAGCTGAGTAAAGGAGGTGAAACAGGAACGCACTTCTACTATCTACCAAAAAA 498
 Db 64 TAGCTCTGAGTAAAGGCGGAAAAACAGGACGCACTTNTACTGAGCTACCAAAAAA 7

RESULT 12

ABL38065

ID ABL38065 standard; cDNA; 605 BP.

XX AC ABL38065;

XX XX

DT 08-APR-2002 (first entry)

XX XX

DE Human colon tumour antigen polynucleotide SEQ ID NO:1654.

XX XX

XX Human; colon cancer; colon tumour antigen; cytostatic; vaccine;

XX colon tumour metastatic antigen; diagnosis; gene; ss.

XX OS Homo sapiens.

XX XX

XX WO200196388-A2.

XX PD

XX 20-DEC-2001.

XX XX

XX 08-JUN-2001; 2001WO-US19557.

XX PF

Mon Nov 24 15:20:15 2003

second nucleic acid molecule comprising any of 15112 nucleotide sequences, appearing as ABX3486-ABX4947, or complements of them. Also included are: (1) a transformed cell having a nucleic acid comprising an LMPD nucleic acid linked to a promoter and a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and (2) determining a level or pattern of a molecule in a bovine cell or tissue comprising: (a) incubating a marker nucleic acid (comprising any of the 15112 nucleic acid sequences or its complement or fragment) with a complementary nucleic acid molecule obtained from the bovine cell or tissue, where hybridisation between the marker nucleic acid and the complementary nucleic acid permits the detection of the molecule; and (b) detecting the level or pattern of the complementary nucleic acid, where the detection of the complementary nucleic acid is predictive of the level or pattern of the molecule. The LMPD nucleic acid is used for determining a level or pattern of a molecule in a bovine cell or tissue. It is useful for genome mapping, gene identification and analysis, cattle breeding, preparation of constructs for use in cattle gene expression, or for genetically improving cattle. The present sequence is one of the 15112 bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The present sequence was not shown in the specification but was obtained in electronic format from the USPTO web site: seqdata.uspto.gov/sequence.html?DocID=20020137139.

Sequence 432 BP; 123 A; 76 C; 78 G; 155 T; 0 other;

Query Match	13.6%;	Score 172.2;	DB 25;	Length 432;
Best Local Similarity	72.2%;	Pred. No. 2.9e-33;		
Matches 294;	Conservative 0;	Mismatches 103;	Indels 10;	Gaps 5;

QY	379	ACTGTAAAGCGCAGCAGCTGATTTCTA-CACCCAGTTACTAGAAACGAGGGGAAGCA	437
DB	432	AAACATAAAGCTGCAACAGCTGATTTTATAACCTAGTTACTAGAAACCTAGGAAGCA	373
QY	438	CTAGTCAGC--TGAGTAAGAGAGGTGAAACAGGAACGACCTTCTACTATCTACCAAAA	495
DB	372	CTTATTAGCTTTGATTAAGCAACATGAAACAGAGGTGCCTTTTACTAACCTACAAAA	313
QY	496	AAATCTCCGAATGCAATTATCAGAAAGATCTTATAGTACAGGTGAGCATATTCCTCGTTA	555
DB	312	AAATTTTCTAATGCATATTAGAAAGATTTTATAATACAAAG-CAGGCATATTGCTCATTA	254
QY	556	AGAGGGGTCTTAAGAAAGACCTTGCTAGCTTAGCACTGTGAGGATGGCCAGTTTA	615
DB	253	AGAATGAGTTCTATAGAAAGCACTTACTAGTTAGCACTATGAGGATGACCAAGTCA	194
QY	616	AATATGAGCTCAACGCC---CCATCTGGGGAGGAGCAGAGGGGGAAGGGGGCTCAAG	671
DB	193	GAGATGATTAAATGCCCAATTTCAGAGGGGTGGCGAGTTTAAGGAAAGCTTAAGAAA	134
QY	672	AGAGACACTGATAGATCGGCCAATTTGTCTACTTG--TTTGACAGAAATTAACCGTTA	729
DB	133	ACACTCACTGATAATACCGACCTTTCTTCATCTACCGCAITTCACAGAAATTAACCTTTT	74
QY	730	AAAGCTTTACCGTGACACTTTTATTCAGTTGATTTACTCCATGTA	776
DB	73	AAAACCTTTACCGGTGATGCTTTTATATAGTTTAAAGCTTTACGTGTA	27

Search completed: November 23, 2003, 11:09:34
Job time : 367.861 secs

Db	841	GTCTGTGTTTACACACTCCACCCAGCAGACCCACGACTAGGAAACAGAAATACTTTCGTTAGA	900
Qy	901	GGCAACACAGGAGCCAGAGTCTTGTTCAAGCCCTGCAGAGCCGGTCAGCTGGTATTTT	960
Db	901	GGCAACACAGGAGCCAGAGTCTTGTTCAAGCCCTGCAGAGCCGGTCAGCTGGTATTTT	960
Qy	961	GAGAACTCACTATGAAATCAAGACGACAGAGCTGTGTACACCCATCGTGACCTACAGTACAA	1020
Db	961	GAGAACTCACTATGAAATCAAGACGACAGAGCTGTGTACACCCATCGTGACCTACAGTACAA	1020
Qy	1021	AGTTACGTAATAGCATGGGCTGATATAGTTACAGTTCGTTATCATGSCAGCGTGTCAATTA	1080
Db	1021	AGTTACGTAATAGCATGGGCTGATATAGTTACAGTTCGTTATCATGSCAGCGTGTCAATTA	1080
Qy	1081	AGGAGGCTGTGCTGTGTGCACACGCTTCGGAGCTACGGGAGGCTCTGCACCCCTGAGCCC	1140
Db	1081	AGGAGGCTGTGCTGTGTGCACACGCTTCGGAGCTACGGGAGGCTCTGCACCCCTGAGCCC	1140
Qy	1141	AGAAGCTGCAGTCTTCTTAAGGACAAAGTCTCTCAACAGCTTAGTGTCTAGTGTTCTCA	1200
Db	1141	AGAAGCTGCAGTCTTCTTAAGGACAAAGTCTCTCAACAGCTTAGTGTCTAGTGTTCTCA	1200
Qy	1201	GCACAAACGCAACTAGTTCACAGGTTATTTGSCAAATCTTAATCTCAGCAAGAATAAGG	1260
Db	1201	GCACAAACGCAACTAGTTCACAGGTTATTTGSCAAATCTTAATCTCAGCAAGAATAAGG	1260
Qy	1261	GATTTT 1266	
Db	1261	GATTTT 1266	
RESULT 2			
AC106124/c			
LOCUS			
DEFINITION	AC106124	217700 bp	DNA linear HTG 10-MAY-2003
ACCESSION	Rattus norvegicus clone CH230-119E11, *** SEQUENCING IN PROGRESS		
VERSION	***, 11 unordered pieces.		
KEYWORDS	AC106124		
SOURCE	HTG; HTGS PHASE3; HTGS DRAFT; HTGS_ENRICHED.		
ORGANISM	Rattus norvegicus (Norway rat)		
REFERENCE	Rattus norvegicus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
	1 (bases 1 to 217700)		
	Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allien,H., Albrooks,S., Amin,A., Anguiano,D., Anyalbechi,V., Aoyagi,M., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Suresh,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Lohliaseg,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,		

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nait, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vexa, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

FEATURES
Source

1. 217700
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-119E11"
BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others
ORIGIN

Query Match 97.0%; Score 1227.4; DB 2; Length 217700;
Best Local Similarity 98.7%; Pred. No. 1.6e-312;
Matches 1237; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY	14	TTTTCAAGTTC	CAAGACATTTTTTTTTTTTTTTTATGATCAAGATTTATTAAGTCA	73
Db	64694	TTTCAAGTTC	CAAGACATTTTTTTTTTTTTTTATGATCAAGATTTATTAAGTCA	64635
QY	74	TACATGCAAA	CACTACTGCTAACTGCTTACAAAGATCAATGTAATAAACAACCTCCACA	133
Db	64634	TACATGCAAA	CACTACTGCTTAACTGCTTACAAAGATCAATGTAATAAACAACCTCCACA	64575
QY	134	TTTGTCAACT	GTCAATTCAGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCACTGTGTT	193
Db	64574	TTTGTCAACT	GTCAATTCAGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCACTGTGTT	64515
QY	194	CTTGCCAGT	CAGTGTAGTTGTACGAACGGGTTAGCACTAGCGCTTGACAAACCTCAC	253
Db	64514	CTTGCCAGT	CAGTGTAGTTGTACGAACGGGTTAGCACTAGCGCTTGACAAACCTCAC	64455
QY	254	AGACCCAAAG	TACCGGAAGCATGTGTCGCGTGGGTAGTCTAGAGGGGGCGGATCA	313
Db	64454	AGACCCAAAG	TACCGGAAGCATGTGTCGCGTGGGTAGTCTAGAGGGGGCGGATCA	64395
QY	314	ATCACATGAC	AGTGTGGTACTCTGGCAACACAGTGTATTCAGAAATATCTAAATAGT	373
Db	64394	ATCACATGAC	AGTGTGGTACTCTGGCAACACAGTGTATTCAGAAATATCTAAATAGT	64335
QY	374	TTAAAACTGT	AAAGCGGACGACGTTTCTACACCCAGTGTACTAGAAAACGAAAGGA	433
Db	64334	TTAAAACTGT	AAAGCGGACGACGTTTCTACACCCAGTGTACTAGAAAACGAAAGGA	64275
QY	434	AGCACTAGT	CAGTGTAGTAAAGGAGTGAACAGCAAGCGACTTCTACTACTACCAA	493
Db	64274	AGCACTAGT	CAGTGTAGTAAAGGAGTGAACAGCAAGCGACTTCTACTACTACCAA	64215
QY	494	AAAAATCTCC	GAATGCATTTATCAGAAAGATCTTATAGTACAGGTACAGATATTTGCTCGT	553
Db	64214	AAAAATCTCC	GAATGCATTTATCAGAAAGATCTTATAGTACAGGTACAGATATTTGCTCGT	64155
QY	554	TAAGAAGGGG	GTCTTAAAGAAAGCACTTGTGAAGTTAGCAACTGTGAGGATGGCAGTT	613

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHLL
Center clone name: CH230-119E11
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185725 bases at least Q40
Consensus quality: 190140 bases at least Q30
Consensus quality: 192842 bases at least Q20
Estimated insert size: 194758; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

gene

```

1543 CGTGACACCTTTTATTCAGTTGAATTAATCTCAATGTACATGTAGTGTAAAGTAGTAATCTCTAC 1484
802 TTCAATATTAGTCAAAATACCTGTCTCTCTTTGATGACGTCGTGTGTTTACACACTCCAC 861
1483 TTCAATATTAGTCAAAATACCTGTCTCTCTTTGATGACGTCGTGTGTTT--CACACTCCAC 1426
862 CCAGCACACCCAGCTAGGAAAGAACTACTCTGTTAGAGGCAACACAGGAGCCAGAGTT 921
1425 CCAGCACACCCAGCTAGGAAAGAACTACTCTGTTAGAGGCAACACAGGAGCCAGAGTT 1366
922 CTGTTCAAGCCCTGCAGAGCCGGTGCAGCTGGTATTTTATAGAGAACTCACTATCAAAATCAA 981
1365 CTGTTCAAGCTGCAAAAGCTAGTGGCTGGTCTTTTATAGAACTCACTATCAAAATCAA 1306
982 AGAGCAGAGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACGTAATGACGATG 1038
1305 AGAGCTGAGCTGTCACACTCATCACTGTGACGTACAGTACAAAGTTACGTAATGACGATG 1246
1039 GCGTGAATAGTTACAGGTCGTTTACATGCGAGCGTGTCTTAAAGAGGCTGTCTGTCTGTC 1098
1245 GCGTGAATAGTTACAGGTCGTTTACATGCGAGCGTGTCTTAAAGAGGCTGTCTGTCTGTC 1186
1099 ACACGGTCTGGAGCTACGGAGGCTCTGACCCCTGAGCCAGGCTGCAAGTCTTCTT 1158
1185 ACACGGTCTGGAGCTCTGGAAGGCTCTGACCCCTGAGCCAGGCTGCAAGTCTTCTT 1126
1159 AGGACAAA---GTCTCTCAACAGCTAGTGTCTGCTTACGTTCTCAGCACACGCACTTA 1215
1125 AGCAACAGAGTCTCTCAAGAGAGTTTGTGCTTGTAGTGTCTCAGCACACGCACTTA 1066
1216 GTTCAACAGTATTTTGGCAATCTTAATCTGAGCAAGAAATAGGGATTT 1266
1065 GTTCAGAGTATTTTGGCAATCTTAATCTGAGCAAGAAATAGGGATTT 1015

```

```

RESULT 5
BC051053/c
LOCUS      2281 bp      mRNA      linear      ROD 14-APR-2003
DEFINITION Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.
ACCESSION BC051053
VERSION    BC051053.1 GI:29835221
KEYWORDS   MGC.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2281)
Direct Submission
Submitted (11-APR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@hgrl.nih.gov
Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, J., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Teurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 108 Row: f Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenesScan gene prediction, Similarity but not identity to protein.

FEATURES	Location/Qualifiers	Source
1..2281		
/organism="Mus musculus"		
/mol_type="mRNA"		
/strain="CZECH II"		
/db_xref="taxon:10090"		
/clone="MGC:58966 IMAGE:5038182"		
/tissue_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."		
/clone_lib="NCI CGAP_Lu29"		
/lab_host="DH10B"		
/note="Vector: pCMV-SPORT6"		
197..775		
/codon_start=1		
/product="Unknown (protein for MGC:58966)"		
/protein_id="AAH51053.1"		
/db_xref="GI:29835222"		
/translation="MQAIKCVVVDGAVGKTCILLISYTNAPFGEYIPTVFDNYSANVWDGKPVNGLMDTAGQEDYDRLPLSYPTDVLICFSLVSPASPNVRKAKYPEVRHPCNPITLLVGTGLDRLDDKDTIEKLKELTPTIPQGLAMAKIGAVKYLECSNLTGRLKTVFDEAIRAVLCPPPPKRRKCLL"		
BASE COUNT 568 a 569 c 504 g 640 t		
ORIGIN		
Query Match 68.6%; Score 868.2; DB 10; Length 2281;		
Best Local Similarity 86.4%; Pred. No. 6.5e-218;		
Matches 1102; Conservative 0; Mismatches 123; Indels 50; Gaps 11;		
QY 35 TTTTITTTTTTTTATGATTTCAAGGATTTTATTAAGTCATACATGCAAAACATCTGCTA 94		
Db 2281 TTTTITTTTTTTTATGATTTCAAGGATTTTATTAAGTCATACATGCAAAACATCTGCTA 2222		
QY 95 ACTGCATTAGCAAAAGATCAATGTAATAAACAATCTGCAATCTGCAACTGTCAATGAAA 154		
Db 2221 ACTGCATTAGCAAAAGATCAATGTAATAAACAATCTGCAATCTGCAACTGTCAATGAAA 2162		
QY 155 AAGATTTGTTCTAGTGTGAAAGGCCCAACACTGTGTTCTTCCAGTGTAGTTAGTTGT 214		
Db 2161 AAGATTTGTTCTAGTGTGAAAGGCCCAACACTGTGTTCTTCCAGTGTAGTTAGTTGT 2102		
QY 215 ACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGG----- 264		
Db 2101 ACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGG----- 2042		
QY 265 -----TACCGGAACCATGTGTCGCGTGGGTGAGGTCTAGAGGGGGCGGATCA 313		
Db 2041 GGCAGAGCCACGACGAGAGCGTCTGTCCACGTGGGCGAGTCTAGA--GGCAGCAATTA 1985		
QY 314 ATCAATGACAGTGTGGTACTCTGCGAACACAGT--GATGTTTCAGATATCTAAATA 371		
Db 1984 GTCAATGACAGTGTGGTACTCTCTGTTAGAGGTTTCAATAGTATCTAAATA 1925		
QY 372 GTTTAAAAACTGTAAAGCCGACGACGTGATTTCTACACCCAGTTACTAGAAACGAAGG 431		
Db 1924 GTTTAAAAACCGTAAAGCTCAACACATGATTTTGCACACCTAGTTGTAGGAACTAAGG 1865		
QY 432 GAAGCACTAGT-CAGCTGAGTAAAGGAGTGAAGCAAGCAAGCACTTCTACTATCTAC 490		
Db 1864 AAGCACTAAGTACTGCTGAGTAAAGTAAAGCGGAAACAGGACGCACTTCTACTAGCTAC 1805		
QY 491 C-AAAAAACTCCGAATGCAATTTATCAGAAAGATCTTATAGTACAGGTGACACATATTCG 549		
Db 1804 CAAAAAAACCCCTTGGTTTATCAGAAAGATCTTATAGTACAGGTGACACATATTCG 1745		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 108 Row: f Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenesScan gene prediction, Similarity but not identity to protein.

QY 550 TCGTTAAGAGGGGGTCTTAAGAAAGACATTTGCTAAGTTAGCAACTGTGAGATGGCC 609
Db 1744 TCGTTAAGAA-GGGGTTCTTAAGAAAGACATTTGCTAAGTTAG-ACATGTGAGGTGGCC 1687
QY 610 AGTTTAAATATGACTCAAGCCCAATCTGGGAGGAGGACAGAGGGGGGAGGGGGCTCA 669
Db 1686 AGTTTAAATATGACTCAAGCCCAATCTGGGAGGAGGACAGAGGGGGGAGGGGGCTCA 1627
QY 670 AG-----AGAGACACTGATAAGATCGGCCATTTGTTCATCTACTCTGTTGACAGA 717
Db 1626 CGAAGGGATGCTCAAGAGACACTGATAAGATTTGGCCATTTGTTCACCTACTCTGTTGACAGA 1567
QY 718 AATTAAACCTTTAAAGACTTTACCGTGACACTTTTATTCAAGTTGAATTAATCTCCATGTAC 777
Db 1566 AATTAAACCTTTAAAGACTTTACCGTGACACTTTTATTCAAGTTGAATTAATCTCCATGTAC 1507
QY 778 ATGTAGTGTAAATTAATCTTACTTCAATATTAGTCAAAATATGCTCTCTCTCTTTGAT 837
Db 1506 AATGTAGTGTAAATTAATCTTACTTCAATATTAGTCAAAATATGCTCTCTCTCTTTGAT 1447
QY 838 GACGTCGTGTTTTCACACTCTCCACCCAGCACACCCAGCAGTACAGCAAAATCTTCGTT 897
Db 1446 CATGTCGTGTTT--CACACTCCACCCAGCACACCCAGCAGTACAGCAAAATCTTCGTT 1389
QY 898 AGAGCAACACAGGAGCCAGAGTTCTGTTCAAGCCTGCAGAAAGCGGTCAGCTGCTATT 957
Db 1388 AGAGCAACACAGGAGCCAGAGTTCTGTTCAAGCCTGCAGAAAGCGGTCAGCTGCTATT 1329
QY 958 TTAGAGACTCACTATGAATCAAGAGAGAGCTGTTTACCCCATC---GTGAGCTACA 1014
Db 1328 TTAGAGAACTCACTATGAATCAAGAGAGCTGTTTACCCCATC---GTGAGCTACA 1269
QY 1015 GTACAAAGTTACGTAATCAGCATGGCTGATAAGTTTACAGTGGCTTACATGGCGAGCGTG 1074
Db 1268 GTACAAAGTTACGTAATCAGCATGGCTGATAAGTTTACAGTGGCTTACATGGCGAGCGTG 1209
QY 1075 TCATTAAGAGGCTGTGCTGTGTACAGCGTCTGGAGCTACGGAGGCTGTGACACCCCT 1134
Db 1208 TCATTAAGAGGCTGTGCTGTGTACAGCGTCTGGAGCTACGGAGGCTGTGACACCCCT 1149
QY 1135 GAGCCAGAGCTGCTGCTTTTAAGGACAAA---GTCTCTCAACAGCTTAGTGCTTAC 1191
Db 1148 GACTCCAGAGCTGCTGCTTTTAGCAACAGAGAGTCTCTCAAGAGAGTTTAGTGCTTAC 1089
QY 1192 GTGTTCTCAGCACACCACTTACTAGTTCACAGAGTATTTTGGCAATCTTAACTGACGA 1251
Db 1088 GTGTTCTCAGCACACCACTTACTAGTTCAGAGGATTTTGGCAATTTTAACTGACGA 1029
QY 1252 AGAATAGGAGTTT 1266
Db 1028 AGAATAGGAGTTT 1014

RESULT 6
AC068493/c
LOCUS
DEFINITION
AC068493
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC068493
Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
AC068493
GI:15148081
HTG; HTGS PHASE1; HTGS DRAFT.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269081)
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 269081)
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,

Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpcgg.org/Sequence/mouse.html>
Contact: hpcgcmendel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-PP - N/A
**Estimated insert size: 268581 - sum-of-contigs
Quality coverage: agarose-PP - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 51730: contig of 51730 bp in length
51731 51750: gap of unknown length
51751 82423: contig of 30673 bp in length
82424 82443: gap of unknown length
82444 118158: contig of 35715 bp in length
118159 118178: gap of unknown length
118179 139441: contig of 21263 bp in length
139442 139461: gap of unknown length
139462 160638: contig of 21177 bp in length
160639 160658: gap of unknown length
160659 177611: contig of 16953 bp in length
177612 177631: gap of unknown length
177632 190991: contig of 13360 bp in length
190992 191011: gap of unknown length
191012 200128: contig of 9117 bp in length
200129 200148: gap of unknown length
200149 211882: contig of 11734 bp in length
211883 211902: gap of unknown length
211903 220773: contig of 8871 bp in length
220774 220793: gap of unknown length
220794 229643: contig of 8850 bp in length
229644 229663: gap of unknown length
229664 238339: contig of 8676 bp in length
238340 238359: gap of unknown length
238360 244848: contig of 6489 bp in length
244849 244868: gap of unknown length
244869 249725: contig of 4857 bp in length
249726 249746: gap of unknown length
249747 252523: contig of 2778 bp in length
252524 252543: gap of unknown length
252544 257520: contig of 4977 bp in length
257521 257540: gap of unknown length
257541 260396: contig of 2856 bp in length
260397 260416: gap of unknown length
260417 262709: contig of 2293 bp in length
262710 262729: gap of unknown length
262730 263934: contig of 1205 bp in length
263935 263954: gap of unknown length
263955 264196: contig of 242 bp in length
264197 264216: gap of unknown length
264217 264229: contig of 413 bp in length


```

Db      1473 CTTTGATCA-----CACACACCCACCCGCGACACACACAGCTA--AACAGAAATT 1426
QY      891 CTTCTTTAGAGCAACACAGGAGCCAGAGTCTTCTTTCAAAGCCTCGAGAGCCGGTCAGC 950
      |||||
Db      1425 CTTCTATTAGAGGAATA-----GCAGTTCCTGTTCAAAATCTCCGCAAAAGCTGCTCAGA 1372
QY      951 TGGTATTTTAGAGAACTCACTATGAATCAAAAGACAGAGCTGTTACACCCATCGTGAGG 1010
Db      1371 AAACTCGCTATGAATCAAAAGACTGATCCAAAGAGCTGAGCTGCTAGCTCACTCCAT 1312
QY      1011 TACAGTACAAAGTTAGTAAATGAGATGAGCTGCTGATAAGTTACAGTGCCTTACATGGCAG 1070
Db      1311 TACAGTACAAATGTTATGTCGGACACAGGCTGCTAACTCACTGCTGAGTTCAATGGCAA 1252
QY      1071 CTTGCAITTAAGAGAGCTGCTGCTGCTACAGGCTCTGG-----GAGCTAGGGAGGG 1123
      |||||
Db      1251 CGCTTCATTCGGGAGGCTGTTCTGCTTTACGCATCTGAGAACTACATAGGAGCAAGTG 1192
QY      1124 TCTGCACC--CCTGAGCCAGAGCTGCACTCTTTTAAAGGACAAAGTCTCTCAACAGCTT 1182
Db      1191 TCTGCACCTCTTAACCTGCGAAGCTACCGTCTTCTCAAGACGAAGTCTTTGCAAGTT 1132
QY      1183 ACTGCTTAGCTGTTCTCAGCAAC---GCAACTTAGTTTCAAGGATATTTGGCAATTC 1239
      |||||
Db      1131 CAGTGTCTGGTGTCTCTCGCACACAACTGAGTGTAGTTTCAAGAGTATTTTGGCAATC 1072
QY      1240 TTAATCTGACGAAGATAGGGG 1261
      |||||
Db      1071 TTAATCTGAACAAGATGGGG 1050
      |||||

RESULT 8
HSA132695/c
LOCUS      HSA132695      28567 bp      DNA      linear      PRI 09-NOV-2000
DEFINITION Homo sapiens rac1 gene.
ACCESSION AJ132695
VERSION AJ132695.5 GI:8574037
KEYWORDS  alternative splicing; Alu; AluJo; AluSg1; AluSp; AluSx; Alu-rich;
          CT-rich; MIR; rac1 gene; rac1 protein; Rac1b protein; repetitive
          sequence.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Matos,P., Skaug,J., Marques,B., Beck,S., Verissimo,F., Gespach,C.,
            Boavida,M.G., Scherer,S.W. and Jordan,P.
  TITLE     Small GTPase Rac1: structure, localization, and expression of the
            human gene
  JOURNAL    Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)
MEDLINE     20517245
PUBMED      11062023
REFERENCE   2 (bases 1 to 28567)
  AUTHORS   Jordan,P.
  TITLE     Direct Submission
  JOURNAL    Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana,
            Laboratorio de Oncobiologia, Instituto Nacional de Saude 'Dr.
            Ricardo Jorge', Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL
  COMMENT    On Jun 20, 2000 this sequence version replaced gi:7248282.
            Related sequence AI638561.
FEATURES
  source     1..28567
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="7p22"
            /clone="H. NH0425A05, RPCI 11"
            1..109
            /rpt_family="AluSg"
            125..422
            /rpt_family="AluSc"
            441..742
            /rpt_family="AluSx"
            4538..4719
            /rpt_family="LTR33A"
            4754..5053
            /rpt_family="AluSg"

```

```

/rpt_family="AluSg"
743..28117
/gene="rac1"
743..1226
/gene="rac1"
911..983
/rpt_family="GC_rich"
936
/gene="rac1"
/note="alternative transcription start site"
1026
/gene="rac1"
/note="alternative transcription start site"
1037
/gene="rac1"
/note="alternative transcription start site"
1053
/gene="rac1"
/note="alternative transcription start site"
1106..1175
/rpt_family="GC_rich"
/rpt_family="GC_rich"
1227..1261,13701..13772,18413..18530,22804..22860,
24268..24330,26010..26169,26458..26588)
/gene="rac1"
/note="alternative"
/codon_start=1
/product="Rac1b protein"
/protein_id="CAA10733.6"
/db_xref="GI:8574039"
/db_xref="SPTREMBL:O95501"
/translation="MQAICVVVGAGVGTCLLSYTTNAPFGEYPTVFDNYSANV
MYDGKPNVLGMDTAGQEDYDLRLPSYPQTGTYGKDTISRGKDKPIADVPLICFS
LVSPASFENVRKAWPEVHRHCPNTPILVGTKLDRDDKDTIEKLEKKLTPITVPQ
GLAMAKEGAVKYLECSALTQRLKATVFEARAVLCPPPVKKRKKCLLL"
Join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MQAICVVVGAGVGTCLLSYTTNAPFGEYPTVFDNYSANV
MYDGKPNVLGMDTAGQEDYDLRLPSYPQTGTYGKDTISRGKDKPIADVPLICFS
LVSPASFENVRKAWPEVHRHCPNTPILVGTKLDRDDKDTIEKLEKKLTPITVPQ
GLAMAKEGAVKYLECSALTQRLKATVFEARAVLCPPPVKKRKKCLLL"
Join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MQAICVVVGAGVGTCLLSYTTNAPFGEYPTVFDNYSANV
MYDGKPNVLGMDTAGQEDYDLRLPSYPQTGTYGKDTISRGKDKPIADVPLICFS
LVSPASFENVRKAWPEVHRHCPNTPILVGTKLDRDDKDTIEKLEKKLTPITVPQ
GLAMAKEGAVKYLECSALTQRLKATVFEARAVLCPPPVKKRKKCLLL"
<1227..1261
/gene="rac1"
/number=1
1262..13700
/gene="rac1"
/number=1
1390..1419
/rpt_family="GC_rich"
1752..1793
/rpt_family="FLAM"
1794..2005
/rpt_family="AluJo"
2140..2171
/rpt_family="Simple_repeat"
2666..2972
/rpt_family="AluSx"
2980..3267
/rpt_family="AluSx"
3268..3289
/rpt_family="(T)n"
3802..4102
/rpt_family="AluJb"
4122..4386
/rpt_family="AluSx"
4538..4719
/rpt_family="LTR33A"
4754..5053
/rpt_family="AluSg"

```



```

repeat_region 5297..5594
/rpt_family="AluSx"
repeat_region 5655..5757
/rpt_family="AluSx"
repeat_region 5758..6062
/rpt_family="AluSx"
repeat_region 6063..6198
/rpt_family="AluSx"
repeat_region 6222..6401
/rpt_family="AluSx"
repeat_region 7162..7238
/rpt_family="AluSg/x"
repeat_region 7348..7367
/rpt_family="L1MC5"
repeat_region 7368..7656
/rpt_family="Simple_repeat"
repeat_region 7890..8183
/rpt_family="AluSx"
repeat_region 8484..8778
/rpt_family="AluSx"
repeat_region 9071..9365
/rpt_family="AluSx"
repeat_region 10275..10583
/rpt_family="AluSp"
repeat_region 10609..10864
/rpt_family="L2"
repeat_region 10970..11276
/rpt_family="AluJo"
repeat_region 11276..11594
/rpt_family="AluDb"
repeat_region 11744..11825
/rpt_family="L2"
repeat_region 11764..11987
/rpt_family="MIR"
repeat_region 12989..13307
/rpt_family="AluSx"
exon 13701..13772
/gene="rac1"
number=2
intron 13773..18412
/gene="rac1"
number=2
repeat_region 13954..14235
/rpt_family="AluSx"
repeat_region 14807..14831
/rpt_family="Simple_repeat"
repeat_region 14836..15128
/rpt_family="AluSx"
repeat_region 15408..15715
/rpt_family="AluSx"
repeat_region 15872..15896
/rpt_family="AT-rich"
repeat_region 15902..16186
/rpt_family="AluSx"
repeat_region 16187..16314
/rpt_family="AluJo"
repeat_region 16413..16585
/rpt_family="Simple_repeat"
repeat_region 16586..16679
/rpt_family="FLAM_A"
repeat_region 16707..16810
/rpt_family="L1M4"
repeat_region 16811..17108
/rpt_family="AluSg"
repeat_region 17109..17345
/rpt_family="L1M4"
repeat_region 17377..17686
/rpt_family="AluSx"
repeat_region 17703..17729
/rpt_family="Simple_repeat"
exon 18413..18530
/gene="rac1"
number=3

```

```

intron 18531..22803
/gene="rac1"
number=3
repeat_region 19287..19436
/rpt_family="MIR"
repeat_region 21161..21184
/rpt_family="Simple_repeat"
repeat_region 21185..21351
/rpt_family="AluSp"
repeat_region 21365..21403
/rpt_family="Alu"
misc_difference 21364^21365
/gene="rac1"
/note="2345 nt insertion sequence in RPCI 11-derived BAC"

Query Match 34.9%; Score 441.8; DB 9; Length 28567;
Best Local Similarity 67.4%; Pred. No. 2.6e-105;
Matches 881; Conservative 0; Mismatches 342; Indels 84; Gaps 15;

QY 10 TTTTTCCTCAAGTTCCTCAAGACATTTTTTTTTTTTTTTTTTTTATGATTCAGGATTTATTAA 69
Db 28148 TTGTTTCATCAAGTCTCAAAAAACACACAGTATTACCAGTCGTATGATTCAGGATTTATTAA 28089
QY 70 GTCATATGTCARAAACATCTGCTAACTGCATTAGCAAAAGATCAATGTAAACACATCC 129
Db 28088 GTCATATGTCARAAACATCTGCTAACTGCATTAGCAAAAGATCAATGTAAACACATCC 28029
QY 130 ACATTCCTGCAACTGTCAATTTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACATG 189
Db 28028 ACATTCCTGCAACTGTCAATTTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACATG 27969
QY 190 TGTTCCTGCGAGTGGATTAGTGTGTACAGAACGGCGTTAGCAGTACAGTACAGTACAGT 247
Db 27968 TATTCCTGCGAGTGGATTAGTGTGTACAGAACGGCGTTAGCAGTACAGTACAGTACAGT 27909
QY 248 CCTCAGACAGCCCAAGG-----TACCGGAACATGTGTCCGCGT 286
Db 27908 CCTCAGACAGCCCAAGGAAACATCAATAGGCAAAAGCGCATCAGGAGCGTGTGTCCGCGT 27849
QY 287 GGGTGAAGTCTAGAGGGGGCGGCATCAATCAGATGACAGTGTGGTACTCTGGCAAGACA 346
Db 27848 GGGCGAGTTAAACA-GGGTCAGTATTGGTCAAGTGACAGTGTGGTAAATCTGGCAAGACA 27790
QY 347 GTGATGTTTCAGAA-----TATCTAAAAATAGTTTAAAAACTCTAAAGC 389
Db 27789 GTGATGTTTAAAGAGTTTCATAGTTTAAAGAAATTAATAAATATTTTAAACATATAAGC 27730
QY 390 CGCAGACAGTCAATTTCTACACCCAGTTACTAGAAAACGAAGGAAGCAGTACAGTACAGT 449
Db 27729 TGCAACACATGATTTTTTACACCTAGTTACTAGAAAACGAAGGAAGCAGTACAGTACAGT 27670
QY 450 GTAAAGGAAGGTGAAACAGGAACGCACTTCTACTATCTACCAAAAACCTCCGATGC 509
Db 27669 GAATA-----AAGTACATGGAAGCACTTTTACTAATCGACAAAAAACCTTTAATGC 27615
QY 510 ATTATCAGAAAGATCTTATAGTACAGTACAGTACAGTATTTGCTCTAGTCAAGAGGGGTCTAT 27556
Db 27614 ATTATCAGAAAGATTTTATAATAACAAG-GAGGCATATTGCTCAGTCAAGAGGGGTCTAT 27556
QY 570 AGAAAAGCACTTGCTAAGTTAGTACAGTGTGAGGATGGCGAGTTTAAATATGACTCAAC 629
Db 27555 AGAAAAGCACTTACTTAAGTTAGTACAGTGTGAGGATGGCGAGTTTAAATATGACTCAAC 27496
QY 630 GCCCATCTGGGAGGACAGCAGGGGGAGGGGCTCAAGAGAGACACTGATAGATC 689
Db 27495 GCCCAATTTGGGAGGACATGGCAGGGTGAAGAGAAAGGAAAGCTTTAAGAAACATTTCC 27436
QY 690 GGCCATTTGCTACTACTGTTTACAGAAATTAACCGTTAAAAAGCTTTACCGTCACAC 749
Db 27435 TGATAATACCAACCTTTCTTTTCATCATCTACTGCTTTCACAGAAATTAACCTTTTAGAG 27376
QY 750 TTTTATTGAGTTGAA---TTACTCCATGTACATGATGATGATAATTTAACTCTACTTCAT 806

```

```
Db 27375 TTTTACCCGTCACACTTTCCTCTGATCAATGATGTA-----AATCTCCACTTCGT 27320
QY 807 A-TTAGTCAAAATAGTCTGTCTCTCTTTGATGACGTCGTGTTTACACACTCCACCCAG 865
Db 27319 ATTGTTGCAAAATAGTCTGTCTCTCTTTGATCA-----CACACACCCACCCCGG 27270
QY 866 CACACCCAGCTAGGAACAGAAATACTCTGTTAGAGGCAACACAGGACGAGTTCGT 925
Db 27269 CACACCCACAGTA--ACAGAAATCTTCAATTAGAGGAATA-----CGAGTCTGTTTC 27218
QY 926 TCAAGCCTCGAAGACCCGTCAGTCGTGATTTTATAGAACTCACTATGAATCAAGAG 985
Db 27217 AAAATCTCGCAAAAGCTGTCAGAAACTCGCTATGAATCACAAGACTGATCAAAAG 27158
QY 986 CAGAGCTGTACACCCATCGTCAGCTACAGTACAGTCAAGTTACGTAATGAGCATGGCTGAT 1045
Db 27157 AGCTGAGCTGCTACGCTCTACTCCATTTACAGTCAATGTTATGTCGGAACACGTCGTCT 27098
QY 1046 AAGTTACAGGTGCTTACATGCGAGCGTGTCTATTAGGAGGCTGTCTGTGTACACAGGT 1105
Db 27097 AACTCACTGCTGAGTTCAATGGCAACGCTTCATTCGGGAGGCTGTTCTGTTTACGCATC 27038
QY 1106 CTGG-----GAGTACCGGAGGGTCTGCACC-CTGAGCCCAAGAGCTGCAGTCTTCT 1157
Db 27037 TGAGAACTACATAGGAGAGCAAGTGTCTGCACCTCCTAACTGCAGAAGCTACCGTCTTCT 26978
QY 1158 TAAGCAAAAGTCTCTCAACAGTTAGTGTCTTACGTTGTTCTCAGCAACAC---GCACTT 1214
Db 26977 CAAAGCAGAGGTCTTTGCAAAAGTTCAGTCTCGGTGTTCTCGGCACAAATGCAAGTGT 26918
QY 1215 AGTTCAAGAGTATTTTGGCAATTTCTTAATCTCAGCAAGAAATAGGGG 1261
Db 26917 AGTTCAGAGGTATTTTGGCACTCTTAATCTCAGCAAGATGGGG 26871

RESULT 9
AC009412/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
ACCESSION AC009412
VERSION AC009412.6 GI:14190769
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
REFERENCE
1 (bases 1 to 212827)
Hou,S., Maupin,R., Haakenson,W., Gregory,S. and Belter,E.
The sequence of Homo sapiens BAC clone RP11-425P5
Unpublished
3 (bases 1 to 212827)
Waterston,R.H.
Direct Submission
Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 212827)
Waterston,R.H.
Direct Submission
Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 212827)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

COMMENT

On May 23, 2001 this sequence version replaced gi:13431187.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

FEATURES

source

1. 212827
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"

repeat_region

/clone="RP11-425P5"
/clone_lib="RP11-11"
1623..1696
/rpt_family="MIR"

repeat_region

1698..2006
/rpt_family="Alu"

repeat_region

2024..2337
/rpt_family="Alu"

repeat_region

3006..3093
/rpt_family="L1"

repeat_region

3453..3552
/rpt_family="L2"

repeat_region

4924..5046
/rpt_family="ERV1"

misc_feature

5318..5693
/note="similar to EST BF996284 (NID:gl2402607)"

repeat_region

5580..5845
/rpt_family="Alu"

repeat_region

5925..6277


```
Db 1542 TA----AATCTCCACTTCGTATTTTGTCAAATACTGTCTTGTCTCTTGATCA----- 1493
QY 846 GTTTCACACTCCACCAGCACACCCAGACTAGAACAAATACTTCGTTCAGAGCAA 905
Db 1492 ----CACACACCCACCCGGCACACCCACAGCTA--AACAGAAATCTTTCATTAGAGGAA 1439
QY 906 CACAGGAGCCAGAGTTCGTTCCAAAGCCTGCAGAGAGCCGGTCAGCTGGTATTTAGAA 965
Db 1438 TAG-----CAGTCTCTGTTCAAATCTCCGAAAAGCTGGTCAGAAAACCTCGCTATGAA 1385
QY 966 CTCCTACTATGAATCAAGAGCAGAGCTGTTCACCCATCGTGACGTACAGTACAAAGTTA 1025
Db 1384 TCACAAAGACTGATCCAAAGAGCTGAGCTGTCTACGCTCACTCCATTAGACATGTTA 1325
QY 1026 CGTAATGAGCATGGCTGATTAAGTTACAGGTGCGTTACATGGCAGCGTGTCAATTAAGAG 1085
Db 1324 TGTCCGGGAACAGCTGCTGCTAACTCACTGCTGAGTTCAATGCAACGCTTCATTCCGGAG 1265
QY 1086 GCTGTGCTGTCTCAAGCTGTGG-----GAGCTACGGGAGGTCTGCACC--CCTGAG 1137
Db 1264 GCTGTCTGCTTTACGATCTGAGAACTACATAGGAGAGAGTGTCTGCACCTCCTAAC 1205
QY 1138 CCCAGAGCTCAGTCTCTTTAGGACAAAGTCTCTCAACAGCTTAGTGTCTTAAGTGTTC 1197
Db 1204 TGCAGAGCTACCGTCTCTCAAGACGAGGTCTTTGCAAGTTCAGTGTCTCGTGTTC 1145
QY 1198 TCACACAAAC---GCACTTAGTTCAAGAGTATTTGGCAATTTCTTAATCTGAGCAAGA 1254
Db 1144 TCGGCACAAACATGAGGTGAGTTTCAAGAGTATTTTGGCACTCTTAATCTGAACAAGA 1085
QY 1255 ATAGGGG 1261
Db 1084 ATGGGG 1078

RESULT 11
AX408985/c
LOCUS AX408985 1232 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1632 from Patent WO0229103.
ACCESSION AX408985
VERSION AX408985.1 GI:21441690
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 1632 11-APR-2002;
JOURNAL GENE LOGIC INC (US)
FEATURES
    source
        1. .1232
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /note="EMBL/GenBank Accession No. D95274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN
Query Match 33.2%; Score 420.2; DB 6; Length 1232;
Best Local Similarity 67.5%; Pred. No. 9.7e-100;
Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;

QY 53 ATTCAAGATTTTAAAGTCATACATGCAAAACATCTGCTAACTGCAATTCAGCAAAAGAT 112
Db 1228 WTCAGAGATTTTAAAGTCATACATGCAAAACATATGCTAATTCATTCAGCAAAAGAT 1169
QY 113 CAATGTAAACACTCCAAATTCGCACTGCAATTCGCAATTCGCAATTCGCAATTCGCAATTC 172
Db 1168 CAATGTAAACACTCCAAATTCGCACTGCAATTCGCAATTCGCAATTCGCAATTCGCAATTC 1109
QY 173 CGAAAGGCCCAACACTGTGTCTTTCGCACTGAGTTAGTTGTACAGAACGGCGTTAGCAC 232
```

```
Db 1108 TCAGGGTCCCACGCTGATTTCTCCAGTAGTTAAAGTTGTACAGAAATCATGTCAGCAC 1049
QY 233 TAGGCG--TTGACAGAACTTCACAGACCCAAAGG-----TACCG 269
Db 1048 TAGCACAGTTTACAGAACCTTCACAGACCCAAAGAACATCAATAGGCAAGACGACTACAG 989
QY 270 GAAGCATGTGTCGCGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCAATCAGACAGTGT 329
Db 988 GAGGCGTGTGTCGCGTGGCGGAGGTAAAGA--GGTCAATATTGTCAGTACAGTGTCT 930
QY 330 GGTACTCTGGCAAGACAGTGTG-----TTTCAGAAATATCTAAAAATAG 372
Db 929 GGTAACTCTGGCAAGACAGTGTATTAAGAAGGTTTCATAGTTTAAAGATATCTAAAAAT 870
QY 373 TTTAAAACTCTAAAGCGCAGCAGTGTATTTCTACACCCAGTTTACTAGAAAACGAAAGG 432
Db 869 TTTAAAACTATTAAGCTGCAACATGATTTTACACCTAGTTTACTAGAAAACCTAAAG 810
QY 433 AAGCACTAGTCAGCTGAGTAAAGGAGGTGAAACAGGAAGCGCACTTCTACTATCTACCA 492
Db 809 AAGCACTATTAGTCTGAATA-----AAGTAACATGGAAGCACTTTTACTTAATCGACA 755
QY 493 AAAAAATCTCCGAATGCAATTATCAGAAGATCTTATAGTACAGGTACAGATATTCGTG 552
Db 754 AAAAAACCTTCTAATGATTTATCAGAAGATTTTATAATAAG--GAGGCATATTGTCTCA 696
QY 553 TTAAGAAGGGGTCTCTAAAGAAAGCACTTGTAAAGTTAGCAACTGTGAGGATGGCAGT 612
Db 695 GTCAAGAGGGTCTTATAGAAAAGCACTTACTAAGTTAGCGACTAACAGAAACCAACAGT 636
QY 613 TTAATAATGCACTCAACGCCCATCTGGGAGGAGCACAGCGGGGAAAGGGGGCTCAAGA 672
Db 635 TTAAGATGAATTAATGCCAATTTGGGAGGCGATGCGAGGTGTAGAGAAAGAAAG 576
QY 673 GAGCACTGATAGATCGGCCATTTGCTACTGTGTTGACAGAAATTAACCGTTAAAA 732
Db 575 CTTAAGAAAAACATTTCTGTATATACCAACCTTTCTTTTCATCATCTACTGCTATGACAG 516
QY 733 AGCTTTACCGGTGACACTTTTATTCAGTTGAA---TTACTCCATGTACAAATGTAGTAA 789
Db 515 AAATTAACCTTTTAGAGTTTTCACCGTGACACTTTCAATCCCTGTACAAATGTAGTGT 457
QY 790 ATTAATCTCTACTTCATA--TTAGTCAAAATACTCTGTCTGTCTCCCTTTGATGACGTCGT 848
Db 456 ---AATCTCCTCTGTTATTTGTCAAAATACTGTCTTTGTCTCTTGTCTCTGATCA----- 409
QY 849 TCACACTCTCCACCCAGCACACCCAGCTAGGAAACAGAAATATCTGTTAGAGGCAACAC 908
Db 408 -CACACACCCACCCCGGCACACCCACAGCTA--AACAGAAATCTTCTCATTAGAGGAAATAG 352
QY 909 AGGAGCCAGAGTTCTGTTCAAAGCCTGCAGAGCCGGTCAGCTGGTATTATTAGAACTC 968
Db 351 -----CAGTTCTGTTCAAAATCTCCGAAAAGCTGTGTCAGAAAACCTCGCTATGAAATCA 298
QY 969 ACTATGAATCAAGAGCAGAGCTGTTTACACCCATCGTGACGTACAGTACAGTCAAAAGTTACGT 1028
Db 297 CAAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTCCATTACAGTACAAATGTTATGT 238
QY 1029 AATGAGCATGGGTGATAGTTACAGTGGTGTACATGGCAGCGTGTCAATTAAGGAGGCT 1088
Db 237 CGGGAACACGTGCTGCTAACTCACTGTTGAGTTCAATGCAACGCTTCATTCCGGGAGGCT 178
QY 1089 GTCTGTGTACACAGGTCTGG-----GAGCTACGGGAGGCTCGCACC--CCTGAGCCC 1140
Db 177 GTTCTGCTTTACCATCTGAGAACTACATAGGAGCAAGTGTCTGCACCTCTCTAACTGC 118
QY 1141 AGAAGCTGCGAGTCTTCTTAAGGCAAAAGTCTCTCAACAGCTTAGTGTCTTACGTTGTTCTCA 1200
Db 117 AGAAGCTACGCTCTCTCAAGAGCAGAGGCTTTGCTAGTTCAGTTCAGTCTCGGTTCTCG 58
QY 1201 GCACAC---GCACCTTAGTTCAGAGGATTTTGGCAATTTTGGCAATTTCTTAATCTGAGCAAGA 1254
```

```

Db      57 GCACAACAATGCAAGTGTAGTTCAGAAAGGTATTTTGGCAACTCTTAATCTGAACAAGA 1
RESULT 12
LOCUS   HUMPO2ST9          1232 bp      mRNA      linear      PRI 11-MAR-1998
DEFINITION Homo sapiens mRNA, clone:PO2ST9.
ACCESSION D25274
VERSION   D25274.1 GI:464185
KEYWORDS  .
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 1232)
AUTHORS   Ishida,Y., Hadano,S., Nagayama,T., Tomiyasu,H., Wakasa,K. and
            Ikeda,J.
TITLE      Isolation and characterization of 21 novel expressed DNA sequences
            from the distal region of human chromosome 4p
JOURNAL   Genomics 22 (2), 302-312 (1994)
MEDLINE   95104839
PUBMED    7545969
REFERENCE 2 (bases 1 to 1232)
AUTHORS   Ishida,Y.
            Direct Submission
            Submitted (11-NOV-1993) Yoshikazu Ishida, Ikeda Genosphere
            project/ERATO/JRDC; Tokai University School of Medicine, Bohseidai,
            Isehara, Kanagawa 259-1193, Japan
            (E-mail:shinji@eng.med.u-tokai.ac.jp, Tel:81-463-91-5095,
            Fax:81-463-91-4993)
            On Mar 25, 1994 this sequence version replaced gi:434743.
COMMENT   Submitted (11-Nov-1993) to DDBJ by:
            Yoshikazu Ishida
            Ikeda Genosphere project/ERATO/JRDC
            Tokai University School of Medicine
            Bohseidai, Isehara
            Kanagawa 259-11
            Japan
            Phone: 0463-91-4056
            Fax: 0463-91-4110.
FEATURES             Location/Qualifiers
     source           1..1232
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="PO2ST9"
                     /tissue_type="brain striatum"
     polyA_signal      1214..1219
     polyA_site        1232
     BASE_COUNT        329 a      240 c      251 g      411 t      1 others
     ORIGIN
Query Match      33.2%; Score 420.2; DB 9; Length 1232;
Best Local Similarity 67.5%; Pred. No. 9,7e-100;
Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;
QY      53 ATTCAAGGATTTATTAGTCATATCATGCAAAACATCTGCTAACTGCATTAGCAAAAGAT 112
Db      1228 WTCAGAGGATTTATTAGTCATATCATGCAAAACATATGCTAATTGCAATTAGCAAAAGAT 1169
QY      113 CAATGTAAAAACACTCCACAATTTCTGCAACTGTCAATTGAAAAAGTTTCTTAGTGGT 172
Db      1168 CAATGTAAAAACACTCCACAATTTCTGCAACTGTCAATTGAAAAAGTTTCTTAGTGGC 1109
QY      173 CGAAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTAGTTGTACAGAACGGCGTTAGCAC 232
Db      1108 TGAAGGGTCCACGCTGTTATTTCTGCGCAGTGAGTTAAGTTGTACAGAACATCTGCAGCAC 1049
QY      233 TAGCGC--TTGACAGAACCTTCACAGACCCCAAGG-----TACCG 269
Db      1048 TAGCAGAGTTTACAGAACCTTCACAGACCCCAAGGACATCAATAGGCAAGGCACTACAG 989

```

RESULT 13
AC102775/c
LOCUS
DEFINITION

AC102775 185245 bp DNA linear HTG 23-MAR-2003
Mus musculus clone RP23-115C10, WORKING DRAFT SEQUENCE, 9 unordered

pieces.
AC102775
VERSION
KEYWORDS
SOURCE
ORGANISM

HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 185245)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-115C10
Unpublished
2 (bases 1 to 185245)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chapel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeAngelis, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,
Lanazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPeeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 185245)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collamore, A., Cook, A., Cooke, P., Corum, B., DeAngelis, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2003 this sequence version replaced gi:22381790.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19647

Center clone name: 115_C10
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 183758 bases at least Q40
Consensus quality: 184136 bases at least Q30
Consensus quality: 184291 bases at least Q20
Insert size: 178000; agarose-fp
Quality coverage: 12.6 in Q20 bases; agarose-fp
Quality coverage: 12.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 53693: contig of 53693 bp in length
53694 53793: gap of 100 bp
53794 56675: contig of 2882 bp in length
56676 56775: gap of 100 bp
56776 60589: contig of 3814 bp in length
60590 60689: gap of 100 bp
60690 67210: contig of 6521 bp in length
67211 67310: gap of 100 bp
67311 82585: contig of 15275 bp in length
82586 82685: gap of 100 bp
82686 103118: contig of 20433 bp in length
103119 103218: gap of 100 bp
103219 132145: contig of 28927 bp in length
132146 132245: gap of 100 bp
132246 182335: contig of 50090 bp in length
182336 182435: gap of 100 bp
182436 185245: contig of 2810 bp in length.

FEATURES
source

1. 185245
/organism="Mus musculus"
/db_type="Genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-115C10"
/clone_lib="RPCI-23 Female Mouse BAC"
1. 53693
/note="assembly_fragment"
clone_end:SP6
vector_side:left
53794..56675
/note="assembly_fragment"
56776..60589
/note="assembly_fragment"
60590..67210
/note="assembly_fragment"
67311..82585
/note="assembly_fragment"
82686..103118
/note="assembly_fragment"
103219..132145
/note="assembly_fragment"
132246..182335
/note="assembly_fragment"
182436..185245
/note="assembly_fragment"
clone_end:T7

BASE COUNT 56870 a 34956 c 35069 g 57550 t 800 others
ORIGIN

Query Match 32.2%; Score 408.2; DB 2; Length 185245;
Best Local Similarity 89.8%; Pred. No. 2.2e-96;
Matches 486; Conservative 0; Mismatches 43; Indels 12; Gaps 4;

```
QY 736 TTTACCCGTGACACTTTTATTCAGTTGAATTAATCAATGTAACAAATGTAAGTAATTAAT 795
Db 607 TTCCCCCATGACACTTTTATTCAGTTGAATTAATCAATGTAACAAATGTAAGTAAT 548
QY 796 CTCTACTTTCATATTAGTCAAAATA---CTGTCTGTCTCTCTTTGATGAGTGTGTTTCA 851
Db 547 CTCTACTTTCATATTAGTCAAAATACTGTCTGTCTCTCTTTGATGAGTGTGTTTCA 490
QY 852 CACACTCCACCCAGCACACCCAGCTAGGAACAGAAATCTCTGTTAGAGGCAACACAGG 911
Db 489 CACACTCCACCCAGCACACCCAGCTAGGAACAGAAATCTCTGTTAGAGGCAACACAGG 430
QY 912 ACCGAGAGTCTGTTCAAGCCCTGAGAGCCGCTCAGCTGCTGTTATTTTGAAGTCACT 971
Db 429 AACCCAGAGTCTGTTCAAGCTGCAAAAGCTAGTGGCTGCTCTTTTGAAGTCACT 370
QY 972 ATGAATCAAGAGCAGAGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACGT 1028
Db 369 ATGAATCAAGAGCAGAGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACGT 310
QY 1029 AATGAGCATGGCTGATTAAGTTACAGTTCGTTTACATGAGGCGTCTCATTAAGGAGCT 1088
Db 309 AATGAGCATGGCTGATTAAGTTACAGTTCGTTTACATGAGGCGTCTCATTAAGGAGCT 250
QY 1089 GTGCTGTGTACACAGCTCTGGAGCTACGGAGGCTGCTGCACCCCTGAGCCACAGAGCTG 1148
Db 249 GTGCTGTGTACACAGCTCTGGAGCTATGGAAGGCTGCTGCACCCCTGAGCTCAGAGAGCTG 190
QY 1149 CAGTCTTCTTAAGCAAAA---GTCTCTCAACAGCTTAGTGTCTGTTCTCAGCACA 1205
Db 189 CAGTCTTCTTAAGCAAAA---GTCTCTCAACAGCTTAGTGTCTGTTCTCAGCACA 130
QY 1206 ACCCACTTAGTTCACAGGATTTTGGCAATTTCTTAATCTGAGCAAGATAGGGGATTT 1265
Db 129 ACCCACTTAGTTCACAGGATTTTGGCAATTTCTTAATCTGAGCAAGATAGGGGATTT 70
QY 1266 T 1266
Db 69 T 69

RESULT 14
AL583884/c
LOCUS
DEFINITION
  Mouse DNA sequence from clone RP23-324B16 on chromosome 15,
  complete sequence.
ACCESSION
  AL583884
VERSION
  AL583884.20 GI:27368255
KEYWORDS
  HTG.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 202342)
REFERENCE
  Smith, M.
  Direct Submission
  Submitted (21-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
  On Dec 23, 2002 this sequence version replaced gi:26788018.
  Sequence from the Mouse Genome Sequencing Consortium whole genome
  shotgun may have been used to confirm this sequence. Sequence data
  from the whole genome shotgun alone has only been used where it has
  a phred quality of at least 30.
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquery@sanger.ac.uk
  -----
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
```

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-324B16 is from the RPI-23 Mouse PAC library from the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6.

FEATURES

Source	Location/Qualifiers
1..202342	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="15"
	/clone="RP23-324B16"
	/clone_lib="RPI-23"
BASE COUNT	60264 a 39265 c 39460 g 63353 t
ORIGIN	
Query Match	32.2%; Score 408.2; DB 10; Length 202342;
Best Local Similarity	89.8%; Pred. No. 2.3e-96;
Matches	486; Conservative 0; Mismatches 43; Indels 12; Gaps 4;
QY 736	TTTACCCGTGACACTTTTATTCAGTTGAATTAATCAATGTAACAAATGTAAGTAATTAAT 795
Db 185051	TTCCCCCATGACACTTTTATTCAGTTGAATTAATCAATGTAACAAATGTAAGTAAT 184992
QY 796	CTCTACTTTCATATTAGTCAAAATA---CTGTCTGTCTCTCTTTGATGAGTGTGTTTCA 851
Db 184991	CTCTACTTTCATATTAGTCAAAATACTGTCTGTCTCTCTTTGATGAGTGTGTTTCA 184934
QY 852	CACACTCCACCCAGCACACCCAGCTAGGAACAGAAATCTCTGTTAGAGGCAACACAGG 911
Db 184933	CACACTCCACCCAGCACACCCAGCTAGGAACAGAAATCTCTGTTAGAGGCAACACAGG 184874
QY 912	AGCCAGAGTTCCTGTTCAAGCCCTGAGAGCCGCTGAGTGTGTTATTTAGAGTCACT 971
Db 184873	AACCAGAGTTCCTGTTCAAGCTGCAAAAGCTAGTGGCTGGTCTTTTGAAGAACTCACT 184814
QY 972	ATGAATCAAGAGCAGAGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACGT 1028
Db 184813	ATGAATCAAGAGCAGAGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACGT 184754
QY 1029	AATGAGCATGGCTGATTAAGTTACAGTTCGTTTACATGAGGCGTCTCATTAAGGAGCT 1088
Db 184753	AATGAGCATGGCTGATTAAGTTACAGTTCGTTTACATGAGGCGTCTCATTAAGGAGCT 184694
QY 1089	GTGCTGTGTACACAGCTCTGGAGCTACGGAGGCTGCTGCACCCCTGAGCCACAGAGCTG 1148
Db 184693	GTGCTGTGTACACAGCTCTGGAGCTATGGAAGGCTGCAACCCCTGAGTCCAGAGCTG 184634
QY 1149	CAGTCTTCTTAAGCAAAA---GTCTCTCAACAGCTTAGTGTCTGTTCTCAGCACA 1205
Db 184633	CAGTCTTCTTAAGCAAAA---GTCTCTCAACAGCTTAGTGTCTGTTCTCAGCACA 184574
QY 1206	ACGCAACTTAGTTCACAGGATTTTGGCAATTTCTTAATCTGAGCAAGATAGGGGATTT 1265
Db 184573	ACGCAACTTAGTTCACAGGATTTTGGCAATTTCTTAATCTGAGCAAGATAGGGGATTT 184514
QY 1266	T 1266

Db 184513 T 184513

RESULT 15

AC105979

LOCUS

DEFINITION

AC105979 192498 bp DNA linear HTG 21-OCT-2002
Mus musculus clone RP24-301E12, WORKING DRAFT SEQUENCE, 9 ordered
pieces.

ACCESSION

AC105979.4 GI:24182089

VERSION

HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS

Mus musculus (house mouse)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Scaurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 192498)

Birken, B., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-301E12

Unpublished

2 (bases 1 to 192498)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,

Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 192498)

Birken, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,

Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,

Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,

Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,

Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,

Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,

Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R.,

Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,

Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Oct 21, 2002 this sequence version replaced gi:20149424.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIDR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: 120199

Center clone name: 301_E_12

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 190070 bases at least Q40

Consensus quality: 191243 bases at least Q30

Consensus quality: 191514 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 191698; sum-of-contigs

Quality coverage: 11.6 in Q20 bases; agarose-fp

Quality coverage: 9.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 9 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 1061: contig of 1061 bp in length

* 1062 1161: gap of 100 bp

* 1162 5714: contig of 4553 bp in length

* 5715 5814: gap of 100 bp

* 5815 11422: contig of 5608 bp in length

* 11423 11522: gap of 100 bp

* 11523 17600: contig of 6078 bp in length

* 17601 17700: gap of 100 bp

* 17701 30157: contig of 12457 bp in length

* 30158 30257: gap of 100 bp

* 30258 38787: contig of 8530 bp in length

* 38788 38887: gap of 100 bp

* 38888 99444: contig of 60557 bp in length

* 99445 99544: gap of 100 bp

* 99545 143263: contig of 43719 bp in length

* 143264 143364: gap of 100 bp

* 143365 192498: contig of 49135 bp in length.

* 192499

* Location/Qualifiers

* 1. 192498

* /organism="Mus musculus"

* /mol_type="genomic DNA"

* /db_xref="taxon:10090"

* /clone="RP24-301E12"

* /clone_lib="RPC1-24 Male Mouse BAC"

* 1. 1061

* /note="assembly_fragment"

* 1162. 5714

* /note="assembly_fragment"

* 5815. 11422

* /note="assembly_fragment"

* 11523. 17600

* /note="assembly_fragment"

* 17701. 30157

* /note="assembly_fragment"

* 30258. 38787

* /note="assembly_fragment"

* 38888. 99444

* /note="assembly_fragment"

* 99545. 143263

* /note="assembly_fragment"

* 143364. 192498

* /note="assembly_fragment"

* clone end:T7

* vector_side:right"

* 53596 a 42471 c 43252 g 52379 t 800 others

* BASE COUNT

* 53596 a 42471 c 43252 g 52379 t 800 others

* 53596 a 42471 c 43252 g 52379 t 800 others

* 53596 a 42471 c 43252 g 52379 t 800 others

Search completed: November 23, 2003, 14:01:54
Job time : 4753.05 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:17:01 ; Search time 289.886 Seconds
(without alignments)
9470.385 Million cell updates/sec

Title: US-09-717-321A-15
Perfect score: 1017
Sequence: 1 cccctattctgtcagatt.....ccttgggtctgtgaggttc 1017

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	1017	100.0	1017	22	AAH22397	Human rac1 gene re
2	1017	100.0	1266	22	AAH22396	Human rac1 contig
3	290.6	28.6	1022	22	AAH22398	Human rac1 gene re
4	290.6	28.6	2051	21	AAF18236	Lung cancer associ
5	290.6	28.6	28567	25	ABT17030	Human MP21 gene Ra
6	285.2	28.0	1232	24	ABK83623	Human cDNA differe
7	285.2	28.0	1232	24	ABN95134	Gene #1632 used to
8	193.2	19.0	605	24	ABL38065	Human colon tumour

C	9	187.2	18.4	605	21	AAA16478	Human colon cancer
10	172.2	16.9	432	25	ABX42447	Bovine EST associa	
11	168	16.5	1318	23	AA565569	DNA encoding novel	
12	164	16.1	2167	24	ABZ35389	Human gene express	
13	126.4	12.4	285	24	ABL71132	Corn tassal-derive	
14	123.2	12.1	416	25	ABX43616	Bovine EST associa	
15	113	11.1	348	22	AAH22399	Human rac1 gene re	
16	111.6	11.0	447	21	AAA89693	Mouse Exo103 nucle	
17	111.2	10.9	422	21	AAA89694	Mouse Rab2 nucleot	
18	103.2	10.1	3740	22	AAO7296	Human reproductive	
19	98.2	9.7	581	24	ABV87661	Human colon cancer	
20	91.6	9.0	565	24	ABV86985	Human colon cancer	
21	87.8	8.6	506	22	AAO80004	Human breast cance	
22	87.6	8.6	470	22	AAO18145	Human breast cance	
23	85.6	8.4	466	22	AAO25750	Human breast cance	
24	85.6	8.4	495	24	ABL37914	Human colon tumour	
25	68	6.7	588	21	AAO16128	Human colon cancer	
26	67	6.6	594	24	ABNG0324	Human cancer relat	
27	62.2	6.1	426	25	ABX42177	Human EST associa	
28	61.4	6.0	378	21	AAC06949	Human secreted pro	
29	59	5.8	404	25	ABX49041	Bovine EST associa	
30	52.4	5.2	162	24	ABX45998	cDNA encoding colo	
31	41.4	4.1	901	25	ABT16468	Human intracellular	
32	39.2	3.9	4081	23	ABL12618	Drosophila melanog	
33	37.4	3.7	13712	24	ABJ33531	Human immune syste	
34	37	3.6	352	22	AAH22400	Human rac1 gene re	
35	37	3.6	2277	25	ABZ58242	Human G-protein co	
36	37	3.6	260209	24	ABZ56564	Human SULF2 genom	
37	36.8	3.6	3014	23	ABL06602	Drosophila melanog	
38	36.8	3.6	3724	23	ABL20520	Drosophila melanog	
39	36.6	3.6	995	22	AAO15320	Human breast cance	
40	36.4	3.6	7892	24	ABX40055	Human chemically p	
41	36	3.5	5098	22	AAO13425	Murine Kit/stem ce	
42	36	3.5	14041	22	AAH48024	Internal control B	
43	36	3.5	16258	24	ABU70376	Chemically treated	
44	36	3.5	16258	24	ABX40038	Human chemically p	
45	36	3.5	611590	21	AAF22303	Arabidopsis thalia	

ALIGNMENTS

RESULT 1
AAH22397
ID AAH22397 standard; DNA; 1017 BP.
XX AC
XX AAH22397;
XX AC
XX DT 22-AUG-2001 (first entry)
XX DE
XX DE Human rac1 gene related nucleotide sequence #1.
KW Identification; toxic; hepatotoxic; differential gene expression;
KW NSAIID; non-steroidal antiinflammatory drug; ds.
XX OS
XX OS Homo sapiens.
XX PN
XX PN WO200138579-A2.
XX PD
XX PD 31-MAY-2001.
XX PF
XX PF 21-NOV-2000; 2000WO-US32049.
XX PR
XX PR 22-NOV-1999; 99US-0166923.
XX PR 18-FEB-2000; 2000US-0183531.
XX PR 20-NOV-2000; 2000US-0717321.
XX (CURA-) CURAGEN CORP.
XX PI
XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX DR
XX DR WPI; 2001-355948/37.

PT Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
XX Disclosure; Page 7-8; 76pp; English.
PS
PS The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent in known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
XX
SQ Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;

Query Match 100.0%; Score 1017; DB 22; Length 1017;
Best Local Similarity 100.0%; Pred. No. 6.9e-308;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCCTATTCTCTCAGATTAAGATTGCGAAATACCTTGTGAACATAGTTGCGTTG 60
DB 1 CCCCTATTCTCTCAGATTAAGATTGCGAAATACCTTGTGAACATAGTTGCGTTG 60
QY 61 CTGAGAACAGTAAGACATTAAGCTTGTGAGAGACTTTGTCTTAAAGAGACTGCACCTTC 120
DB 61 CTGAGAACAGTAAGACATTAAGCTTGTGAGAGACTTTGTCTTAAAGAGACTGCACCTTC 120
QY 121 TGGGCTCAGGGTGCAGACCCCTCCCGTAGCTCCAGACCGTGTGACACAGCAGCCTCC 180
DB 121 TGGGCTCAGGGTGCAGACCCCTCCCGTAGCTCCAGACCGTGTGACACAGCAGCCTCC 180
QY 181 TTAATGACACGCTGCCATGTAAGCAGCTGTAATATCAGCCCATGCTCATTACCTAAC 240
DB 181 TTAATGACACGCTGCCATGTAAGCAGCTGTAATATCAGCCCATGCTCATTACCTAAC 240
QY 241 TTGTGACTGTACGTACAGATGGGTGTACAGCTCTGCTCTTTGATTTTATGAGTTCT 300
DB 241 TTGTGACTGTACGTACAGATGGGTGTACAGCTCTGCTCTTTGATTTTATGAGTTCT 300
QY 301 CTTAAATACAGCTGACCGGCTTCTGAGCTTTGAAACAGAACTCTGGCTCCTGTGTTGC 360
DB 301 CTTAAATACAGCTGACCGGCTTCTGAGCTTTGAAACAGAACTCTGGCTCCTGTGTTGC 360
QY 361 CTCTACGAGATATTCTGTCTAGTCTGCTGGGTGCTGGGTGAGTGTGAAACACGA 420
DB 361 CTCTACGAGATATTCTGTCTAGTCTGCTGGGTGCTGGGTGAGTGTGAAACACGA 420
QY 421 CGTCTATCAAGGAGACAGACAGTATTTTGACTAATATGAGTACAGATTAATTTACACTA 480
DB 421 CGTCTATCAAGGAGACAGACAGTATTTTGACTAATATGAGTACAGATTAATTTACACTA 480
QY 481 CATTTGATGAGTAAATCACTGAATAAAGTGCACGGTAAGACTTTTAAACGGTT 540
DB 481 CATTTGATGAGTAAATCACTGAATAAAGTGCACGGTAAGACTTTTAAACGGTT 540
QY 541 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGCTCTCTTGAGCCCC 600
DB 541 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGCTCTCTTGAGCCCC 600
QY 601 CTTTCCCTCTGCTGCTCCCTCCAGATGGGGGCTTGAGTCCATATTTAAACTGGCCATCC 660
DB 601 CTTTCCCTCTGCTGCTCCCTCCAGATGGGGGCTTGAGTCCATATTTAAACTGGCCATCC 660

QY 661 TCACAGTTGCTAACTTAGCAAGTGCTTTTCTTTAGGACCCCTCTTAAACGAGCAATAG 720
DB 661 TCACAGTTGCTAACTTAGCAAGTGCTTTTCTTTAGGACCCCTCTTAAACGAGCAATAG 720
QY 721 TCTGACCTGTACTATAAGATCTTTCTGATATGCAATCGGAGATTTTTTTGGTAGATAGT 780
DB 721 TCTGACCTGTACTATAAGATCTTTCTGATATGCAATCGGAGATTTTTTTGGTAGATAGT 780
QY 781 AGAAGTGGTTCCTGCTTTTACCTTCTTACTCAGCTGACTAGTGCCTTCCTTCTGTTT 840
DB 781 AGAAGTGGTTCCTGCTTTTACCTTCTTACTCAGCTGACTAGTGCCTTCCTTCTGTTT 840
QY 841 CTAGTAACCTGGGTGAGAAATCACGCTGCTGGGCTTTACAGTTTAAACTATTTAGAT 900
DB 841 CTAGTAACCTGGGTGAGAAATCACGCTGCTGGGCTTTACAGTTTAAACTATTTAGAT 900
QY 901 ATTCTGAAACATCACTGCTTGTGCCAGAGTACCAACACTGTATGATGATGCGGCC 960
DB 901 ATTCTGAAACATCACTGCTTGTGCCAGAGTACCAACACTGTATGATGATGCGGCC 960
QY 961 CTCTAGACCTTACCCACGCGGACATGCTTCCGGTACCTTTGGGTCTGTGAGGTTTC 1017
DB 961 CTCTAGACCTTACCCACGCGGACATGCTTCCGGTACCTTTGGGTCTGTGAGGTTTC 1017

RESULT 2

AAH22396/c
ID AAH22396 standard; DNA; 1266 BP.

XX AAH22396;

DT 22-AUG-2001 (first entry)

XX Human rac1 contig SEQ ID NO:2.

XX Identification; toxic; hepatotoxic; differential gene expression;

KW NSAID; non-steroidal antiinflammatory drug; ds.

XX Homo sapiens.

XX WO200138579-A2.

XX 31-MAY-2001.

XX 21-NOV-2000; 2000NO-US32049.

XX 22-NOV-1999; 99US-0166923.

XX 18-FEB-2000; 2000US-0183531.

XX 20-NOV-2000; 2000US-0717321.

XX (CURA-) CURAGEN CORP.

XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX WPI; 2001-355948/37.

XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -

XX Disclosure; Page 7; 76pp; English.

XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a

CC hepatotoxic agent in known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX
 SQ Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;
 Query Match 100.0%; Score 1017; DB 22; Length 1266;
 Best Local Similarity 100.0%; Pred. No. 7.7e-308;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTCGCAAAATACCTTGTGAACCTAAGTTGGTCTG 60
 DB 1261 CCCCTATTCTTGGCTCAGATTAAAGATTCGCAAAATACCTTGTGAACCTAAGTTGGTCTG 1202
 QY 61 CTGAGAACACGTAAGCACTAAGCTGTGTGAGAGACTTTGTCTTAAGAGACTGCAGCTTC 120
 DB 1201 CTGAGAACACGTAAGCACTAAGCTGTGTGAGAGACTTTGTCTTAAGAGACTGCAGCTTC 1142
 QY 121 TGGGCTCAGGGGTGACAGACCTCCCGTAGCTCCAGACCGGTGTGACAGACGACGCTCC 180
 DB 1141 TGGGCTCAGGGGTGACAGACCTCCCGTAGCTCCAGACCGGTGTGACAGACGACGCTCC 1082
 QY 181 TTAATGACGCTCCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCAATTACGTAAC 240
 DB 1081 TTAATGACGCTCCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCAATTACGTAAC 1022
 QY 241 TTTGTACTGTACGTCAGATGGGTGTAAACGCTCTGCTCTTTGATTTCATAGTGAGTTCT 300
 DB 1021 TTTGTACTGTACGTCAGATGGGTGTAAACGCTCTGCTCTTTGATTTCATAGTGAGTTCT 962
 QY 301 CTAATAATACAGCTGACCGCTTCTGAGGCTTTGACAGAACTCTGCGCTCTGTGTGC 360
 DB 961 CTAATAATACAGCTGACCGCTTCTGAGGCTTTGACAGAACTCTGCGCTCTGTGTGC 902
 QY 361 CTCTAAGCAAGTATTCGTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTAACACGA 420
 DB 901 CTCTAAGCAAGTATTCGTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTAACACGA 842
 QY 421 CGTCATCAAGGAGACAGACGATTTTGTACTAATATGAAGTAGAGATTAAATTACACTA 480
 DB 841 CGTCATCAAGGAGACAGACGATTTTGTACTAATATGAAGTAGAGATTAAATTACACTA 782
 QY 481 CATTTACATGGAGTAATTCACGTATATAAGTGTACCGGTAAAGCTTTTAAACGTT 540
 DB 781 CATTTACATGGAGTAATTCACGTATATAAGTGTACCGGTAAAGCTTTTAAACGTT 722
 QY 541 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTTGAGCCCC 600
 DB 721 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTTGAGCCCC 662
 QY 601 CCTTCCCGCTGTCTCCCTCCAGATGGGGCTTGAGTCCATATTTAACTGGCCATCC 660
 DB 661 CCTTCCCGCTGTCTCCCTCCAGATGGGGCTTGAGTCCATATTTAACTGGCCATCC 602
 QY 661 TCACAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTCTTAACAGCAATATG 720
 DB 601 TCACAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTCTTAACAGCAATATG 542
 QY 721 TGTGACCTGTACTATAAGATCTTTCTGTGATAATGCAATTCGAGATTTTGTGTAGATAGT 780
 DB 541 TGTGACCTGTACTATAAGATCTTTCTGTGATAATGCAATTCGAGATTTTGTGTAGATAGT 482
 QY 781 AGAAGTGCCTTCTGTTTTCACCTCTCTTACTCAGCTAGTGTCTTCCCTTCTGTTT 840
 DB 481 AGAAGTGCCTTCTGTTTTCACCTCTCTTACTCAGCTAGTGTCTTCCCTTCTGTTT 422
 QY 841 CTAGTAACCTGGGTGTAGAAATCACGTGCTGCGGCTTTACAGTTTAACTATTTTAGAT 900
 DB 421 CTAGTAACCTGGGTGTAGAAATCACGTGCTGCGGCTTTACAGTTTAACTATTTTAGAT 362
 QY 901 ATTCTGAACATCACTGTCTTGCAGAGTACCAACACTGTCTCAITGTGATTCGCCGCC 960

DB 361 ATTCTGAACATCACTGTCTTCCAGAGTACCAACACTGTCAATGTGATGATGCCGCC 302
 QY 961 CTCTAGACTCACCACCGGACACATCTTCCGTAACCTTTGGCTGTGTGAGTTTC 1017
 DB 301 CTCTAGACTCACCACCGGACACATCTTCCGTAACCTTTGGCTGTGTGAGTTTC 245

RESULT 3
 AAH22398
 ID AAH22398 standard; DNA; 1022 BP.
 XX AAH22398;
 XX 22-AUG-2001 (first entry)
 XX Human rac1 gene related nucleotide sequence #2.
 XX Identification; toxic; hepatotoxic; differential gene expression;
 XX NSAID; non-steroidal antiinflammatory drug; ds.
 XX Homo sapiens.
 XX WO200138579-A2.
 XX 31-MAY-2001.
 XX 21-NOV-2000; 2000WO-US32049.
 XX 22-NOV-1999; 99US-0166923.
 XX 18-FEB-2000; 2000US-0183531.
 XX 20-NOV-2000; 2000US-0717321.
 XX (CURA-) CURAGEN CORP.
 XX Gould-Rothberg BE, Dipippo VA, Ramseh TW, Gerwein RW;
 XX WPI; 2001-355948/37.

Screening hepatotoxic agent comprises contacting test cell population
 expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 with reference population and identifying difference in expression
 levels -

Disclosure; Page 7-8; 76pp; English.

The present invention describes a method of screening a test agent for
 hepatotoxicity. The method comprises: (a) providing a test cell
 population comprising a cell capable of expressing one or more nucleic
 acid sequences selected from the group consisting of RISKMARKER 1-8
 and INJURYMARKER 1-10; (b) contacting the test cell population with a
 test agent; (c) measuring expression of one or more of the nucleic
 acid sequences in the test cell population; (d) comparing the
 expression of the nucleic acid sequence in the test cell population to
 the expression of the nucleic acid sequence in a reference cell
 population comprising at least one cell whose exposure status to a
 hepatotoxic agent is known; and (e) identifying a difference in
 expression levels of the RISKMARKER or INJURYMARKER sequences, if
 present, in the test cell population and reference cell population.
 The method is useful for identifying a hepatotoxic agent. The present
 sequence is given in the exemplification of the present invention.

Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;

Query Match 28.6%; Score 290.6; DB 22; Length 1022;

Best Local Similarity 64.5%; Pred. No. 3.1e-80;

Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTCGCAAAATACCTTGTGAACCTAAGTTGGC---GTT 57

DB 1 CCCCCATTTCTTGTTCAGATTAGAGTTGCCAAAATACCTTCTGAACCTACACTGSCATTGTT 60

QY 58 GTGCTGAGAACACGTAAGCACTTAAGCACTTTGTGTGAGAGACTTTGTCTTAAAGACACTGCAGC 117

Db	61	GTGCCGAGAAACCCGAGCACTGAACTTTGCAAGACCTTCGTCTTTGAGAAGACGGTAGC	120
Qy	118	TTCTGGGCTCAGG--GGTGCAGACCTCCCGTAGC-----TCCAGACCGGTGTGACACA	169
Db	121	TTCTGCAGTTAGGAGGTCAGACACTTGTCTCTCTATGTAGTTCTCAGATCGTAAAGCA	180
Qy	170	GCACAGGCTCTTAATGACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCT	229
Db	181	GAACAGGCTCCCGAATGAAGGTTGCCATTGAACTCACACGTAGTTAGCAGCAGCTGTT	240
Qy	230	CATTACGTAACTTTGTACTGTACGTACGATCGGTGTAAACAGCTCTGCTCTTTGATTTC	289
Db	241	CCCGACATAACATTGTACTGTAATGGAGTGAGCGTAGCAGCTCAGCTCTTTGGATCAGTC	300
Qy	290	TAGTGAAGTTCTCTAAAATPACAGCTGACCGGCTTCTGCAGGCTTTGAAACAGAACTCTCGC	349
Db	301	TTTGTGATTTTCATAGCGAGTTTCTGACCAGCTTTTGGGAGATTTTGAACAGAACTG--	358
Qy	350	TCCTGTGTGTGCTCTAAACGAAGTATCTGTTCTTCTAGTCTGGGTGTCTGGGTGAGTGT	409
Db	359	----CTATTTCTCTCTAAATGAAGAAATCTGTT--TAGCTGTGGGTGTCCGGGTGGGGTGT	412
Qy	410	GTGAAACACGAGCTCATCAAAAGGAGACAGACAGTATTTTGAC--TAATATGAAGTAGAGAT	468
Db	413	GT-----GTGATCAAAGGACAAGACAGTATTTTGACAAATACGAAGTCGAG--	460
Qy	469	TAATTTACACTACATGTGACATGGAGTAA---TTCAACTGAATAAAGTGTACCGGTAA	525
Db	461	--ATTTACACTACATGTGACAAAGGAATGAAGTGTACCGGTAAAAACTCTAAAGGTTA	518
Qy	526	AGCTTTTAAACGGTTAATTTCTGTCAAACAGTAGATGACAAATGGCCGATCTTATCAGTG	585
Db	519	ATTCTGTCAAATGCAGTAGATGATGAAGAAGGTGGTATTAATCAGAAATGTTTCT	578
Qy	586	TCTCTTTGAGCCCCCTTCCCGCTGTCTCTCCCGAGATGGGCGGTGAGTCCATAT	645
Db	579	TAAGCTTTTCTCTCTTACACCTGCCATGCTCCCAAATTTGGGCATTTAATTCATCT	638
Qy	646	TTAACTGGCATCTCACAGTGTCTAACTTAGCAAGTCTTTCTTTTAGGACCCCTTC	705
Db	639	TTAACTGGTGTCTGTGTAGTCGCTTAACTTAGTAGTCTTTTCTTATAGAACCCCTTC	698
Qy	706	TTAAGCAGCAATATGTCTGACCTGACTATAAGATCTTTCTGATAATGATTCGGAGATT	765
Db	699	TCAGCTGAGCAATATGCCT--CCTGTATTATAAAATCTTCTGATAATGATTAGAAGTT	757
Qy	766	TTTTTGTGTAGATAGTAGAGTGGTTCCTGTTTTCACCTTCTTTTACTCAGCTGACTAGT	825
Db	758	TTTTTGTGATAGTAGTAAAGTGTCTTCCATGTATC-----TTTATTACAGACTAATAGT	812
Qy	826	GCTTCCCTTCGTTTTCTAGTAACTGGGTGTAGAAATCAGCTGTGGGCTTTTACAGTTTT	885
Db	813	GCTTTTCCITAGTTTTCTAGTAAGTGTAAAAATCATGTGTTCAGCTTTATAGTTTT	872
Qy	886	TAAACTATTTTTAGATA-----TTCTGAAACATCACTGCTCTGCCAGAG	928
Db	873	TAAAAATTTTAGATAAATTTTAAACTATGAACCTTTCTTAACATCACTGCTTTCGCAGAT	932
Qy	929	TACCAACACTGTCATGTGATTTGATGGCGCCCTCTAGACCTCACCCACGCGACACATG	988
Db	933	TACCGACACTGTCACCTTGACCAATCTG--ACCTCTTTACCTCGCCACGCGACACAG	991
Qy	989	CTTCCGGTA	997
Db	992	CCTCTCTGTA	1000

RESULT 4
AAF18236
ID AAF18236 standard; DNA; 2051 BP
XX
AC AAF18236;

XX	14-MAR-2001 (first entry)
DT	Lung cancer associated polynucleotide sequence SEQ ID 255.
XX	
DE	
XX	
XX	Human; lung cancer associated protein; neuroprotective; cytostatic;
KW	cardioactive; immunomodulatory; muscular active; vulnerary;
KW	gastrointestinal; nephrotropic; antiinfective; gynecological;
KW	antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW	proliferative disorder; wound healing; infectious disease; ds.
XX	
OS	Homo sapiens.
XX	
XX	WO2000055180-A2.
XX	
XX	21-SEP-2000.
XX	
XX	08-MAR-2000; 2000WO-US05918.
XX	
XX	12-MAR-1999; 99US-0124270.
XX	
XX	(HUMA-) HUMAN GENOME SCI INC.
FA	(ROSE/) ROSEN C A.
PA	
XX	
XX	Ruben SM;
PI	
XX	
XX	WPI; 2000-587514/55.
DR	P-PSDB; AAB58360.
DR	
XX	
XX	Lung cancer associated gene sequences, referred to as lung cancer
PT	antigens, useful for treatment, prevention, and diagnosis of disorders
PT	such as lung cancer
XX	
XX	Claim 1; Page 716-717; 1425pp; English.
PS	
XX	
CC	Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC	associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC	associated proteins and polynucleotide sequences, their agonists, and
CC	antagonists may have neuroprotective; cytostatic; cardioactive;
CC	immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC	general; nephrotropic; antiinfective; gynecological; or antibacterial
CC	activity. The invention also includes antibodies specific for the
CC	protein or polynucleotide sequences. The lung cancer associated
CC	Polynucleotide sequences may be used for detection of lung cancer,
CC	chromosome identification, as chromosome markers, and for numerous other
CC	diagnostic or research purposes. The proteins may be used to treat
CC	disorders such as neural, immune, muscular, reproductive,
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC	disorders. The proteins may also be used in the treatment of wounds and
CC	infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC	peptide AAB58349 are used in the course of the invention for the
CC	identification and characterisation of the polynucleotide and protein
CC	sequences.
XX	
SO	Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;

	Query Match	28.6%;	Score	290.6;	DB	21;	Length	2051;
	Best Local Similarity	64.5%;	Pred. No.	4.5e-80;				
	Matches	664;	Conservative	0;	Mismatches	304;	Indels	61; Gaps
								13;
Qy	1	CCCCATTCTTGTCCAGATTAAAGNATGCCAAATAACCTTTGTAACCTAAGTTGC--GGTT	57					
Dd								
	755	CCCCCACTCTGTTTCAGATTAAAGATTGCCAAATAACCTTTGTAACCTAAGTTGTT	814					
Qy	58	GTGCTGGAACAACGTAAGCAGCTAAGCTGTTTGAGAGACTTTTGTCTTTAAGAAGACTGCAGC	117					
Dd								
	815	GTGCCGAGAACACCGAGCACTGAACCTTTGTCAAAGACCCTTCGCTCTTTGAGAAGACGGTAGC	874					
Qy	118	TTCGGGCTCAGG--GGTGCAGACCCCTCCCGTAGC-----TCCAGACCGCTGTGCACACA	169					
Dd								
	875	TTCGTGAGTTAGGAGGTGCAGACACTTGTCTCTCATGTAGTCTTCAGATCGGTAAGCA	934					
Qy	170	GCACAGCGCTCCCTTAATGACACGCTGCCACTGTAAACGACCTGTAACTTATGAGCCCATGCT	229					

Current Month	20 68:	Score	290 6:	DB 21:	Length	2051:
---------------	--------	-------	--------	--------	--------	-------

Query Match	28.6%;	Score: 230.0; Z=227;	Long-matching
Best Local Similarity	64.5%;	Pred. No. 4.5e-80;	
Mismatches	0;	Mismatches 304;	Indels 61; Gaps 13;
Matches	664;	Conservative	

1 CCCCTATTCTTGCTCAGATTAGAAATTGCCAAAATACCTTGTGACTAAGTGC--GTT 57

b
755 CCCCATTCTTGTTTCAGATTAAAGAGTTGCCAAAATACCTTCTGAACCTACACTGCATTGTT 814

58 GTGCTGAGAACACGTAAGCCTAAGCCTGTTGAGAGAGACTTGTCTCTTAGAGACTGCAGC 117

b
815 GTGGCGAGACACCGAGCACTGAACTTTGCAAGACCTTCGTCTTTGAGAGACGGTAGC 874

118 TTCTGGGCTCAGG-GGTCAGACCTCCCGTAGC-----TCCGAGACCGGTGTGACACA 169

b
875 TTCTGCAGTTAGGAGGTGCAGACACTTGCTCTCCTATGTAGTTCTCAGATGGCTAAAGCA 934

170 GCACAGCCTCCTTAATGACACGCTGCCATGTAACGCACCTGTAACCTATCAGCCCATGCT 229

AAFT 8236;

```
Db 935 GAAAGAGCTCCCGAATGAAGGTTGCCATTGAATCACCAGTGAGTTAGCAGCAGTGTT 994
QY 230 CATTACGTAACCTTTGACTGTACGTACAGTCAGATGGGTGTAACAGCTCTGCTCTTTGATTCA 289
Db 995 CCGACATACCAATGTACTGTAAATGGAGTGAAGTAGCAGCTCAGCTCTTTGGATCAGTC 1054
QY 290 TAGTGAGTTCTCTAAATACACAGCTGACCGGCTTCTGCAGGCTTTGAACAGAACTCTGGC 349
Db 1055 TTTGTGATTCATAGCAGTTTCTGACCAGCTTTTGGGAGATTTTGAACAGAACTG-- 1112
QY 350 TCCGTGTGCTCTAACGAAGTATTCCTGTTCTAGTGTGGGTGCTGGGTGGAGTGT 409
Db 1113 ---CTATTTCTCTAATGAAGATTTCTGTT--TAGCTGTGGGTGTCGGGTGGGTGT 1166
QY 410 GTGAACACAGCCTCATCAAGGAGACAGACAGTATTTTGC--TAATATGAAGTAGAGAT 468
Db 1167 GT-----GTGATCAAGGACAAAGACAGTATTTTGACAAATACGAAGTGGAG-- 1214
QY 469 TAATTTACACTACATTCATGATGAGTAA---TTCACTGAATAAAGTGTACGGGTAA 525
Db 1215 --ATTACACTACATTCATGAGGAATGAAGTGTACGGGTAAAACTCTAAAGGTTA 1272
QY 526 AGCTTTTAAAGGTAAATTTCTGTCAACAGTAGATGACAAATGCCCGATCTTATCAGTG 585
Db 1273 ATTTCTGTCAATGCAGTAGATGATGAAGAAAGGTTGGTATTATCAGAAATGTTTCT 1332
QY 586 TCTCTCTGAGCCCTTCCCTCTGCTGCTCCCTCCAGATGGGGTGTGAGTCCATAT 645
Db 1333 TAAGCTTTTCTTCTTACACTGCTGCTGCTCCCTCCAAATGGGCATTTAATTCATCT 1392
QY 646 TTAACCTGGCCATCCCTCACAGTTGCTAACTAGCAAGTGTCTTTCTTTAGACCCCTTC 705
Db 1393 TTAACCTGGTGTCTGTTAGTGTCTTAACTAGTGTCTTTCTTATGAACCCCTTC 1452
QY 706 TTAACGAGCAATGCTGACCTGTACTATPAAGATCTTTCTGATAATGCAATTCGGAGATT 765
Db 1453 TGACTGAGCAATATGCT--CTTGATTAATAAATCTTTCTGATAATGCAATTAAGGTT 1511
QY 766 TTTTGGTAGATAGTAGAAGTGTGCTGCTGTTTACCTTCCCTTACTACGTGACTAGT 825
Db 1512 TTTTGTGCAATTAATAAAGTGTCTTCCATGTTAC-----TTTATTCAGAGCTTAATAAGT 1566
QY 826 GCTTCCCTTCTGTTTCTAGTAATCGGTGAGAAATCAGTGTGCGGCTTACAGTTT 885
Db 1567 GCCTTCTTAGTTTCTAGTAATCAGTGTGTAATAATCATGTGTGCGCTTTATAGTTT 1626
QY 886 TAAACTATTTTATGATA-----TTCTGAACATCACTGTCTTGGCCAGAG 928
Db 1627 TAAATAATTTTATGATAATTTTAAACTATGAACCTTCTTAAACATCACTGTCTGCGAGAT 1686
QY 929 TACCAACCTGTCATGTGATGATGCGGCCCTCTAGACCTCACCACGGGACACATG 988
Db 1687 TACCACTGTCATGTGACCAATAGT--ACCTCTTTACCTGCGCCAGCGGACACAG 1745
QY 989 CTTCCGGTA 997
Db 1746 CTTCTGTGA 1754
```

RESULT 5

ABT17030

ID ABT17030 standard; DNA; 28567 BP.

XX

AC ABT17030;

XX

DT 03-APR-2003 (first entry)

XX

DE Human MP21 gene Rac1 SEQ ID No 4.

XX

KW Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;

XX

cell proliferation disorder; MP21; gene; ds.

XX

OS Homo sapiens.

XX WO2003006990-A1.

XX

XX 23-JAN-2003.

XX

XX 10-JUL-2002; 2002WO-US21549.

XX

XX 12-JUL-2001; 2001US-305017P.

XX 10-OCT-2001; 2001US-328491P.

XX 15-FEB-2002; 2002US-357452P.

XX

XX (EXEL-) EXELIXIS INC.

XX

XX Friedman L, Plowman GD, Belvin M, Li D, Funke RP;

XX WPI; 2003-221779/21.

XX P-PSDB; ABJ19756.

XX

Identifying candidate p21 pathway modulator, by contacting an assay system having modifiers of p21 polypeptide or gene with a test agent to provide a reference activity in system and detecting test agent-biased activity -

Examples; Page 56-72; 199pp; English.

The invention relates to a novel method for identifying a candidate p21 pathway modulating agent. The novel method comprises contacting an assay system, comprising a purified MP21 polypeptide (modifier of p21) or nucleic acid, with a test agent under conditions, so that but for the presence of a test agent, the assay system provides a reference activity and detection of test agent-biased activity of the assay system. The novel method of the invention is useful for identifying a candidate p21 pathway modulating agent. The invention also includes a method for modulating the p21 pathway of a cell, and a method for diagnosing a disease e.g. cancer in a patient. The identified modulators are useful in diagnosis, therapy and pharmaceutical development. The modulators are useful in a variety of diagnostic and therapeutic applications including angiogenic, apoptotic and cell proliferation disorders. This polynucleotide sequence represents a gene encoding an MP21 protein of the invention.

Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other;

Query Match 28.6%; Score 290.6; DB 25; Length 28567;

Best Local Similarity 64.5%; Pred. No. 2e-79;

Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATCTTCTGCTAGATTAAGAAATGCGAAATACCTTGTGAACCTAAGTTGC---GTT 57

Db 26871 CCCCCATCTTCTTTCAGATTAAGAGTTGCCAAATACCTTCTGAACCTACACTGCATTGTT 26930

QY 58 GTGCTGAGAACACGTAAGCACTTAAGCTGTGTGAGACTTTGTCTTAAAGAACTGCAGC 117

Db 26931 GTGCCGAGAACCCGAGCACTGAACCTTTGCAAGAACCTTCGCTTTGAGAAGACGCTAGC 26990

QY 118 TTCTGGGCTCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCCAGACCGGTGACACA 169

Db 26991 TTCTGCAGTTAGAGGTGCAGACACTTGCTCTCTTATGTAGTTCTCAGATGCGTAAGCA 27050

QY 170 GCACAGCCTCTTAATGACACGCTGCCATGTGAACGACCTGTAACTTATCAGCCCATGCT 229

Db 27051 GAACAGCCTCCGGAATGAAGCGTTGCCATTGAACCTACCAGTGAAGTGTAGCAGCAGTGT 27110

QY 230 CATTACGTAACCTTTGACTGTACGTACGTACAGATGGGTGTACAGCTCTGCTCTTTGATTCA 289

Db 27111 CCGGACATAACAATTTACTGTACTTAATGGAGTGAAGCTCAGCTCAGCTCTTTGGATCAGTC 27170

QY 290 TAGTGAGTTCTCTAAATACCAACAGCTGACCGGCTTCTGCAGGCTTTTGAACAGAACTCTGCG 349

Db 27171 TTTTGATTTTCATAGCAGATTTTCTGACACAGCTTTTGGGAGATTTTGAACAGAACTG-- 27228

QY 350 TCCTGTGTTGCTCTAACGAAGTATTTCTGTTCTAGTCTGGGTGCTGCTGGGTGAGTGT 409

Db 27229 -----CTATTTCTCTTAAAGAAATTCGTTT--TAGCTGTGGTGTCCGGTGGGTGT 27282
Qy 410 GTGAACACAGCAGTCATCAAAAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 27283 GT-----GTGATCAAGGACAAAGACAGTATTTTGACAAATACGAAGTGGAG-- 27330
Qy 469 TAATTTACATACATTTGACATGAGTAA-----TTCACTGAATAAAGTGTACCGGTAA 525
Db 27331 --ATTACATACATTTGACAGGAATGAAAGTGTACCGGTAAATACTTAAAGGTTA 27388
Qy 526 AGCTTTTAAACGGTTAATTTCTGTCACACAGTAGATGACAAATGCCATCTTATCAGTG 585
Db 27389 ATTCTGTCAAAATGCAGTAGATGATGAAGAAGGTGCTTATATCAGAAATGTTTTCT 27448
Qy 586 TCTCTCTTTGAGCCCCCTTCCCTCTGTCCTCCCGAGATGGGGCGTTGAGTCCATAT 645
Db 27449 TAAGCTTTTCTTCTTTACACCTGCCATGCTCCCAAAATTTGGCATTTAATTCATCT 27508
Qy 646 TTAACCTGGCCATCTCAGTGTCTAAGTGTCTAAGTGTCTTTCTTTAGGACCCCTTC 705
Db 27509 TTAACCTGGTGTCTGTTAGTGTCTAAGTGTCTTTCTTTATAGAACCCCTTC 27568
Qy 706 TTAACGACATATCTGTGACCTGTACTATATAGATCTTTCTGATATGCAATTCGGAGATT 765
Db 27569 TGACTGAGCAATATGCTTCTTTGATATATAAATCTTTCTGATATGCAATTAGAGTT 27627
Qy 766 TTTTGTGTAGATGATGAGAGTGTCTTCTGTTTTCACCTTCCTTTACTCAGTGTACTAGT 825
Db 27628 TTTTGTGTAGATGATGAGAGTGTCTTCTGTTTTCACCTTCCTTTACTCAGTGTACTAGT 27682
Qy 826 GTTCTCTCTGTTTCTAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 885
Db 27683 GTTCTCTCTGTTTCTAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 27742
Qy 886 TAAACTATTTTATGATA-----TTCTGAACATCATCTCTTCTGTCAGAG 928
Db 27743 TAAATATTTTATGATAATTTCTTAAATATGAACTTTCTTAAATCATCTCTTCTGTCAGAT 27802
Qy 929 TACCACACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 988
Db 27803 TACCACACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 27861
Qy 989 CTTCCGGTA 997
Db 27862 CCTCCTGTA 27870

RESULT 6

ABK83623

ID ABK83623 standard; cDNA; 1232 BP.

XX

AC ABK83623;

XX

DT 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #194.

DE

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;

KW viral infection; parasitic infection; protozoal infection;

KW fungal infection; sterile inflammatory disease; psoriasis;

KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;

KW cardiac reperfusion injury; renal reperfusion injury; ARDS;

KW adult respiratory distress syndrome; inflammatory bowel disease;

KW Crohn's disease; ulcerative colitis; periodontal disease;

KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

OS

XX WO200228999-A2.

PN

XX 11-APR-2002.

PD

XX

PF 03-OCT-2001; 2001WO-US0821.

XX 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

PA Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

PI WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression

DR of genes associated with granulocyte activation, which serves as

XX diagnostic markers that is useful for monitoring disease states and

XX drug toxicity

XX Claim 1; SEQ ID No 194; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation

XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by

XX DNA chip analysis as given in the specification, and comparing

XX the expression level to an expression level in an unactivated

XX GC, where differential expression of Gs is indicative of GCA.

XX Also included are modulating (M2) GA by contacting GC with an agent

XX that alters the expression of at least one gene in Gs. (2) screening (M3)

XX for an agent capable of modulating GCA or an inflammation (especially

XX chronic) in a tissue, an allergic response in a subject, exposure of a

XX subject to a pathogen or sterile inflammatory disease using the

XX gene expression profile; (3) detecting (M4) an inflammation (especially

XX chronic) in a tissue, an allergic response in a subject, exposure of a

XX subject to a pathogen or sterile inflammatory disease, by detecting the

XX level of expression in a sample of the tissue of gene(s) from Gs, where

XX the level of expression of the gene is indicative of inflammation;

XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,

XX an allergic response in a subject, exposure of a subject to a pathogen

XX or sterile inflammatory disease, by contacting a tissue having

XX inflammation with an agent that modulates the expression of gene(s)

XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for

XX modulating GA; M3 is useful for screening an agent capable of modulating

XX GCA preferably in an inflammation in a tissue; M4 is useful for

XX detecting an inflammation (especially chronic) in a tissue, an allergic

XX response in a subject, exposure of a subject to a pathogen or sterile

XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal

XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion syndrome,

XX reperfusion injury, ARDS, adult respiratory distress syndrome,

XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,

XX periodontal disease; also bacterial infection, viral infection,

XX parasitic infection, protozoal infection, fungal infection and M5 is

XX useful for treating one of the above conditions. The present

XX sequence represents a gene differentially expressed in granulocytes.

XX Note: The sequence data for this patent did not form part

XX of the printed specification, but was obtained in electronic

XX format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX


```
Db 181 CTCCGAATGAAGCTTGCATTGAACACACAGTGTACAGACAGTGTCCGACA 240
QY 237 TAACCTTGTACTGTACGTACAGATGGGTGTACAGCTCTCTTGTGATTCATAGTGAG 296
Db 241 TAACATTTGTACTGTAAATGGAGTGTAGAGCTCAGCTCTTTGGATCAGTCTTTGTGA 300
QY 297 TTCTCTAAATACCAGCTGACCGGCTTCTGAGGCTTTGAACAGACACTGCTCTCTG 356
Db 301 TTTCATAGCGAGTTTCTGACCGCTTTTGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGGCTCTAACGAGTATTTCTGTCTAGTGTGGTGTCTGGGTGTAGTGTGGAAC 416
Db 355 TTTCCTCTAATGAAGATTTCTGTT--TAGCTGTGGTGTGCGGGTGGGTGT----- 407
QY 417 ACGAGCTCATCAAAAGGACAGACAGTATTTTGAC-TAATATGAAGTAGAGATTAATTA 475
Db 408 -----GTGATCAAGAGCAAGACAGATTTTGTCAAAATACGAAGTGGAG----ATT 458
QY 476 CACTACATTTGACATGGAGT---AATTCAACTGAATAAAGTGTCAAGGTAAAGCTTTT 532
Db 459 CACTACATTTGACAGGAATGAAGTGTCAAGGTAAAGTCTTAAGGTTAATTTCTG 518
QY 533 TAACGGTTAATTTCTGTCACAAACAGATAGACAAATGCGCATCTTATCAGTGTCTCT 592
Db 519 TCAATATCAGTAGATGATGAAGAAAGTGTGTTATATCAGGAATGTTTCTTAAGCTT 578
QY 593 TGAGCCCTTCTCCCTGTCTCCCTCCAGATGGGGTGTAGTCCATATTAAGTAACT 652
Db 579 TTCTCTTCTTACACCTGCGATGCCCTCCCAATGGGCATTTAATCATCTTTAACT 638
QY 653 GGCCATCTCACAAGTCTTCACTTAGCAAGTGTCTTTCTTTAGGACCCCTCTCTTACGA 712
Db 639 GGTGTTCTGTAGTCTGCTAATCTAGTAGTGTCTTTCTTATAGAACCCCTTCTGACTGA 698
QY 713 GCAATATGCTGACCTGTACTATAAGATCTTTCTGATAATGCAATTCGGAGATTTTGTG 772
Db 699 GCAATATGCT-CCTTGTATTATAAAATCTTTCTGATAATGCAATGCAAGTGTTTTGT 757
QY 773 TAGATAGTGAAGTGGTCTCTCTTTCTACCTTCTTCTACCTGCTGACTAGTGTCTCC 832
Db 758 CGATTAGTAAAGTGTCTTCCATGTTAC-----TTTATTGAGAGCTAATAAGTCTTCC 812
QY 833 TTGCTTTCTAGTAACTGGGTGTAGAAATCAGCTGTCTGGGCTTTACAGTTTAACTA 892
Db 813 TTAGTTTCTAGTAACTAGTGTAAATCATGTGTGAGCTTTATAGTTTTTAAATA 872
QY 893 TTTTAGATA-----TTCTGAACATCACTGTCTTCCAGAGTACCAAC 935
Db 873 TTTTAGATAATCTTAAACTATGAACCTTCTTAACATCACTGTCTTCCAGATTAACGAC 932
QY 936 ACTGTCATGTGATTGATGCGCCCTCTAGACCTCACCACCGGACACATGCTTCCGG 995
Db 933 ACTGTCATGTGACCAATACTG-ACCCTCTTTACCTCGCCACCGGACACAGCTCTCTG 991
QY 996 TA 997
Db 992 TA 993
```

RESULT 7

ABN95134

ID ABN95134 standard; DNA; 1232 BP.

XX AC

XX ABN95134;

XX DT

13-AUG-2002 (first entry)

XX DE

Gene #1632 used to diagnose liver cancer.

XX KW

Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;

KW metastatic liver tumour; cytostatic; expression profile; disease state;

KW disease progression; drug toxicity; drug efficacy; drug metabolism.

```
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US30589.
XX PR 02-OCT-2000; 2000US-237054P.
XX (GENE-) GENE LOGIC INC.
XX PA Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX PI WPI; 2002-426119/45.
XX DR Diagnosing and detecting the progression of liver cancer,
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,
XX PT involves detecting the level of expression of two or more genes in a
XX liver tissue sample -
XX PS Claim 1; SEQ ID NO 1632; 298pp; English.
XX CC The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumour in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN95134-ABN97455 in a
XX CC tissue sample. The method of the invention has hepatotropic, and
XX CC cytostatic activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
```

Query Match

28.0%; Score 285.2; DB 24; Length 1232;

Best Local Similarity 64.4%; Pred. No. 1.7e-78;

Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;

```
QY 8 TCTTGCTCAGATTAAAGATTGCCAAATACCTTGTGAACCTAGTTGC---GTTGCTGA 64
Db 1 TCTTGCTCAGATTAAAGATTGCCAAATACCTTGTGAACCTAGTTGC---GTTGCTGA 60
QY 65 GAACACGTAAAGCACTAAGCTGTTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTCTGGG 124
Db 61 GAACACCGAGCACTGAACCTTAGCAAGACCTTCGCTTTTGAGAGACGTTAGCTTCTGCA 120
QY 125 CTCAGG-GGTGAGACCCCTCCCGTAGC-----TCCACAGCCGTGTGACAGACAGC 176
Db 121 GTTAGAGGTGCAGACACTTGTCTCTCTATGTAGTTCTCAGATGCTTAAAGCAGACAGC 180
QY 177 CTCCTTAATGACCGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATAGC 236
Db 181 CTCCTGAATGAACGCTTGCCTTGAACCTCACCAGTGTAGTGTAGCAGCAGCTGTTCCGACA 240
QY 237 TAACTTTGTACTGTACTGCTCAGCTCAGATGGGTGTAAACGCTCTGCTCTTTGATTCATAGTAG 296
Db 241 TAACTTTGTACTGTAAATGGAGTGTAGCAGCTAGCTCTTTGGATCAGCTTTGTGA 300
QY 297 TTCTCTAAATACCAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTGCTCTCTG 356
Db 301 TTTCATAGCGAGTTTCTGACCGCTTTTGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGCTCTTAAACGAGTATTTCTGCTAGTGTGGTGTGCTGGGTGTAGTGTGGAAC 416
Db 355 TTTCCTCTAATGAAGATTTCTGTT--TAGCTGTGGTGTGCGGGTGGGTGT----- 407
```

QY 417 ACAGCGTCATCAAGGACAGACAGTATTTTGAC-TAATATGAAGTAGAGATTAATTA 475
 Db 408 -----GTGATCAAGGACAAAGACAGTATTTTGACAAAATACGAAGTGGAG-----ATTTA 458
 QY 476 CACTACATGTACATGAGT---AATTCACCTGAATAAAGTGTACAGGGTAAAGCTTTT 532
 Db 459 CACTACATGTACAGGAATGAAGTGTACAGGGTAAAGTGTCTAAAGGTTAATTTCTG 518
 QY 533 TAACGGTTAATTTCTGTCAACAGTAGATGACAAATGCGCGATCTTATCAGTGTCTCT 592
 Db 519 TCAATGAGTAGATGATGAAGAAGGTTGGTATTATCAGAAATGTTTCTTAAGCTT 578
 QY 593 TGAGCCCCCTTCCCTGTGTCCTCCAGATGGGGGTGAGTCCATATTTAACT 652
 Db 579 TTCCTTTCTTTACACCTGCCATGCTCCCAATTTGGCATTTAATTCATCTTTAAACT 638
 QY 653 GGCCATCCTCAGTGTGCTAACTAGCAAGTGTCTTTTGTAGGACCCCTTCTTAACGA 712
 Db 639 GGTGTTCTGTGAGTGTAACTTAGTAGTGTCTTTCTTATAGAACCCCTTCTGACTGA 698
 QY 713 GCAATATGTCGACCTGTACTATAAGATCTTTCTGATAATGCAATTCGGAGATTTTGTG 772
 Db 699 GCAATATGCTT-CCTGTATTAATAAATCTTCTGATAATGCAATTCGGAGTGTGTTTGT 757
 QY 773 TAGATAGTAGAGTGTGCTGTTTTCACCTTCTTACTCAGCTGACTAGTCTTCCC 832
 Db 758 CGAATAGTAAAGTGTCTTCCATGTATC-----TTTATTCAGAGCTAATAAGTGTCTTCC 812
 QY 833 TTCGTTTCTAGTAACTGGTGTGAATCAGCTGTGGGCTTTTACAGTCTTTTAACTA 892
 Db 813 TTAGTGTCTAGTAACTAGGTGTAATAAATCATGTGTGAGCTTTATAGTGTTTAAATA 872
 QY 893 TTTTAGATA-----TTCTGAAACATCACTGTCTTGGCAGAGTACCAAC 935
 Db 873 TTTTAGATAATCTTAACTATGAACCTTCTTAACATCACTGTCTTGGCAGATTACCGAC 932
 QY 936 ACTGTCATGTATGATGATGCGCCCTCTAGACCTCACCACGCGGACATGCTTCCG 995
 Db 933 ACTGTCATGTGACCAATAGT-ACCCTCTTACCTGCGCCACGCGGACATGCTTCCG 991
 QY 996 TA 997
 Db 992 TA 993

RESULT 8
 ABL38065/c
 ID ABL38065 standard; cDNA; 605 BP.
 XX ABL38065;
 AC
 XX
 DT 08-APR-2002 (first entry)
 DE Human colon tumour antigen polynucleotide SEQ ID NO:1654.
 XX
 XX Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 KW colon tumour metastatic antigen; diagnosis; gene; ss.
 XX Homo sapiens.
 OS
 XX WO200196388-A2.
 FN
 XX 20-DEC-2001.
 PD
 XX 08-JUN-2001; 2001WO-US18557.
 PF
 XX 09-JUN-2000; 2000US-2108992.
 PR 20-FEB-2001; 2001US-270216F.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 PI Jiang Y, Harlocker SL, Secrlist H;

XX WPI; 2002-114514/15.
 DR Novel isolated colon tumor polynucleotide differentially expressed in
 XX colon tumor or colon metastatic tumor and polypeptides encoded by them,
 PT useful for inhibiting development of cancer in patient -
 PT
 XX Claim 1; SEQ ID 1654; 105pp; English.
 PS
 XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
 CC which were isolated from human colon tumour and colon metastatic tumour
 CC cDNA libraries. (II) have cytostatic activity and can be used in vaccine
 CC production. (I) can be used for stimulating and/or expanding T cells
 CC specific for a tumour protein on contact with the T cells. They are also
 CC useful for inhibiting the development of cancer in a patient. (I) can be
 CC used as probes or primers for nucleic acid hybridisation, for preparing
 CC mutant species primers, or primers for use in genetic constructions. (I)
 CC can be used in the diagnosis of a colon tumour.
 XX
 SQ Sequence 605 BP; 219 A; 109 C; 117 G; 156 T; 4 other;
 Query Match 19.0%; Score 193.2; DB 24; Length 605;
 Best Local Similarity 70.5%; Pred. No. 7.7e-50;
 Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
 QY 505 GAATAAAAGTGTACCGGTAAAGCTTTTAAACGGTTAATTTCTGTCAA--CAGTAGATG 562
 Db 592 GAATGAAAGTGTACCGGTAAAGCTTTAAAGGTTAATTTCTGTCAAATGCGTAGATG 533
 QY 563 ACAAT-----GGCCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCC 608
 Db 532 ATGAAAGAAAGTGTGTTATATCAGAAATGTTTCTTAAGCTTTTCTTCTTCTTACAC 473
 QY 609 CTGCTGTCTCCCTCCAGATGGGGCTGTAGTCCATTTAACTTAACTGCGCATCTCACAGTT 668
 Db 472 CTGCCATGCTCCCAAAATGGGCATTTAATTCATCTTTAACTGTTGTTCTGTGTAGTC 413
 QY 669 GCTAATCTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAAACGAGCAATATGTTCACT 728
 Db 412 GCTAATCTAGTAAAGTGTCTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT 354
 QY 729 GTACTATAAGATCTTTCTGATTAATGCAATCGGAGATTTTTTGGTAGATAGTAGATGC 788
 Db 353 GTATTATAAAATCTTTCTGATAATGCAATAGAGGTTTCTTTCGATTAGTAAAGTGC 294
 QY 789 GTTCTGTGTTTTCACCTTCTTCTTACTCAGTGTCTTCTGCTTCTGCTTCTTCTAGTAAC 848
 Db 293 TTTCAGTTTAC-----TTTATTCAGAGCTAATAAGTGTCTTCTTCTTCTAGTAAC 239
 QY 849 TGGGTGTAGAAATCAGTGTCTGCGGCTTTTACAGTTTAACTATTTTATGATA----- 901
 Db 238 TAGGTGTAAATCATGTGTGTCAGCTTTATAGTCTTTTAAATATTTTATGATAATTTTA 179
 QY 902 -----TTCTGAAACATCACTGTCTTGCAGAGTACCACACATGTCATCTGATTGA 951
 Db 178 AACTATGAACCTTCTTAAACATCACTGTCTTGCAGATTTACCGACATGGCATTGACCAA 119
 QY 952 TGCGGCCCTCTCTAGACCTCACCACGCGGACATGCTTCCGGTA 997
 Db 118 TACTG-ACCCTCTTTTACCTGCGCCACGCGGACACACGCTTCTCTGTA 74
 RESULT 9
 ABL16478/c
 ID ABL16478 standard; DNA; 605 BP.
 XX ABL16478;
 AC
 XX
 DT 14-JUN-2000 (first entry)
 DE Human colon cancer differentially expressed nucleotide sequence #483.
 KW Colon cancer; detect; differential expression; human; treatment;
 XX

CC improving cattle. The present sequence is one of the 15112 bovine
 CC IMFED ESR (expressed sequence tag) nucleic acids.
 CC Note: The present sequence was not shown in the specification but
 CC was obtained in electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
 XX
 SQ Sequence 432 BP; 123 A; 76 C; 78 G; 155 T; 0 other;
 Query Match 16.9%; Score 172.2; DB 25; Length 432;
 Best Local Similarity 72.2%; Pred. No. 2.4e-43;
 Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;
 QY 486 TACATGGAGTAATCACTGATTAAGTGTACCGGTAAAGCTTTTAAAGCTTAATTT 545
 DB 27 TACACGTAAACATTAATACTATTAAAGCATCACGGTAAAGTTTAAAGCTTAATTT 86
 QY 546 GTGTCAAAA--CAGTAGATGACAAATGGCGGATCTTTATCAGTGTCTCTTGAGCCCC--- 600
 DB 87 GTGTCAAAATGGGTAGATGAAGAAAGGTGCGTATTATCAGTGAGTGTCTTTAAAGCTTT 146
 QY 601 -CCTTCCCGCTGCTGCTCCCTCCAGATGGGCGGTGATGTCATATTAACTGGCCATC 659
 DB 147 TCCCTTAAACCTCCCAACCCCTTCTGAAATGGGCATTTAATCATCTCTGACCTGCTATC 206
 QY 660 CTCACAGTTGCTAACTTAGCAAGTGTCTTTCTTTAGGACCCCTCTTAAAGCAATAT 719
 DB 207 CTCATAGTGTCTAACTTAGTAAAGTGTCTTTCTTTAGAACTCATCTTTAATGAGCAATAT 266
 QY 720 GTCTGACCTGTACTATAGATCTTTCTGATAATGCAATTCGGAGATTTTTTGGTAGATAG 779
 DB 267 GCCT-CCITGTATTATAAATCTTCTGAATATGCAATAGAAAATTTTTTGTAGTTAG 325
 QY 780 TAGAGTGGGTCCTGTTTTCACCTTCCTTACTC--AGCTGACTAGTCTCCCTTCGT 837
 DB 326 TAAAAGTGACCTCCTGTTTTCATGTTGCTTTTAAAGCTTAATAGTGTCTTCTCTAGT 385
 QY 838 TTTCTAGTAACCT-GGCTGAGAAATCACGCTGTGGGGCTTTACAGTT 883
 DB 386 TTTCTAGTAACCTAGTTTATAAATCACGTTGTGAGCTTTATGGTT 432

RESULT 11
 AAS65569
 ID AAS65569 standard; cDNA; 1318 BP.
 AC AAS65569;
 XX
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #1373.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG01382.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 1; SEQ ID No 1373; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
 Query Match 16.5%; Score 168; DB 23; Length 1318;
 Best Local Similarity 58.1%; Pred. No. 9.4e-42;
 Matches 615; Conservative 0; Mismatches 365; Indels 78; Gaps 15;
 QY 1 CCCCTATTCTTCTGCTCAGATTAAGAATGCCAAATACCTTGTGAAGTATGTC---GTT 57
 DB 39 CCCCCATTTCTTCTCAGATTAAGAGTTGCCAAATACCTTGTGAAGTATGTCATGTT 98
 QY 58 GTGCTGAGAACACGTAAGCACTAAGCTGTGTGAGAGACTTTGTC-CTTAAGAGACTGCG 116
 DB 99 GTGCCGAGAACACCGAGCACTGAATTTGCAAGACCTTCGTCTTTGAGAGACCGTAG 158
 QY 117 CTTCTGGGCTCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCACAGCCGTGTGACA 167
 DB 159 CTTCTCAGTTAGGAGGTGCAGACACTTTGCTCTCTCTATGTAGTTCTCAGATCGCTAAAG 218
 QY 168 CAGCAGAGCTCTTAAATGACAGCGTGCATGTAGCAGCCTGTACTTA-TCAGCCCAT 226
 DB 219 CAGAACAGCCTCCCGAATGAAGCGTTGCCAATGAACCTCAGTGTAGTTACGAGCAGCGT 278
 QY 227 GCTCATTAGTAACCTTTGCTACTGTAGCTCAGATGGGTGTAAAGCTCTGCTTTGATT 286
 DB 279 GTTCCGACATACATTTGCTGTATATGA--GTGAGCGTAGCAGCTCAGCTCTTTGGA 336
 QY 287 TCATAGTGAAGTTCTTAAATACCAGCTGACCGGCTTCTGAGGCTTTTGAACAGACTCT 346
 DB 337 TCAGTCTTTGTGATTTCTAGCGAGTTTCTGACCAGCTTTTCGGGAGATTTTGAACAG 396
 QY 347 GGCTCCTGTGTTGCCTCTAACGAGTATTCTGTCTCTAGTCGCGGTGTCTGGTGGAG 406
 DB 397 AACTGGCTATTTCTCTTAATGAAGGAATTTCTGGTTAGCTGTGGGTGTGCCGGTGGG 456
 QY 407 TGTGTCAAAACACGACGCTCATCAAGGAGACAGACAGTATTTTGTACTAATATGAGTAGAG 466
 DB 457 GGTGTGGTGGTGG-----ATCAAGGACAAAGACAGTAAATTT---TGGACAAATACCG 508
 QY 467 ATTAATTTTACCTACATTTGTACAGTAAATTCACCTGAATAAAAGTGTACCGGTAAA 526
 DB 509 AGTGGGAGGATTTTACCACCTACCATTTGGTACAGGGAATGAAAGTGTGTACCGGGTAAAA 569
 QY 527 GCCTTTTAAAGGTTAATTTCTGTCAACAGTATGACAAATGCCGATCTTTTACGTGT 586
 DB 569 CTCCTAAAAGGGTTAATTTCTGTGTCATGTCAGTAGATGAATGAAGAAAGTTTGTAT 628

OS Zea mays.
XX US2001051335-A1.
XX 13-DEC-2001.
XX 16-APR-1999; 99US-0294093.
XX 21-APR-1998; 98US-082567P.
XX (LALG/) LALGUDI R V.
XX (ITOL/) ITO L Y.
XX (SHER/) SHERMAN B K.
XX Lalgudi RV, Ito LY, Sherman BK;
XX WPI; 2002-163647/21.
XX Novel purified corn tassel-derived polynucleotide useful for
PT determining altered gene expression, to recover regulatory elements and
PT to follow inheritance of desirable characteristics through hybrid
PT breeding programs -
XX Claim 1; SEQ ID 506; 201pp; English.
XX The present sequence describes a purified corn tassel-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76833. The cdps sequences
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (I) are also useful
CC in the evaluation, and alteration of desired characteristics associated
CC with growth and development, disease resistance, environmental
CC adaptability, quality and yield, and as molecular markers for studying
CC inheritance of multigene traits in a plant breeding program. (I) can be
CC used to produce a tassel-specific profile of gene transcription, a
CC transcript image, to clone regulatory elements for use in transformation
CC vectors, to express a polypeptide, to identify, isolate or extend
CC identical or related corn tassel nucleic acid sequences from DNA
CC libraries, in nucleic acid hybridisation or amplification technologies,
CC as query sequences to determine homology of known sequences, as probe
CC for use in Southern or Northern hybridisation, and to identify the
CC presence of and/or to determine the degree of similarity between two
CC (or more) nucleic acid sequences.
XX Sequence 285 BP; 79 A; 46 C; 44 G; 115 T; 1 other;
SQ
Query Match 12.4%; Score 126.4; DB 24; Length 285;
Best Local Similarity 72.7%; Pred. No. 4.4e-29;
Matches 192; Conservative 0; Mismatches 66; Indels 6; Gaps 2;
QY 658 TCCTCAGCTGCTACTAGCAAGTGCCTTTCTTTAGGACCCCTCTTAACGACCAAT 717
Db 5 TTCTGTAGTGCCTACTAGTAGAGTCTTTCTTATAGAACCCCTCTGACTGACCAAT 64
QY 718 ATGCTGACCTGCTACTATAGATCTTTCTGATATGCAATCGAGATTTTGGTAGAT 777
Db 65 ATGCCCT-CCTGTGATTATATAAATCTTTCTGATAATGCAATGAGGTTTTTGTGCAAT 123
QY 778 AGTAGAAGTGCCTGCTGCTGTTTTCACCTCTCCTTACTCAGCTGCTAGTCTCCCTTCGT 837
Db 124 AGTAAAGTGCCTTCCATGTTAC-----TTTATTCAGAGCTATATAGCTTTCCTTAGT 178
QY 838 TTCTAGTAACCTGGGTGTAGAAATCAGCTGCTGCGGCTTTACAGTTTAACTATTTTA 897
Db 179 TTCTAGTAACCTAGGTGTAAATAATCATGTGTTGACGCTTTATAGTTTAAATAATTTAG 238
QY 898 GATATCTGAAACATCACTGCTT 921
Db 239 ATAATCTAAACATATGAACCTCT 262

RESULT 14

ABX43616/c
ID ABX43616 standard; cDNA; 416 BP.

XX

AC ABX43616;

XX

DT 21-FEB-2003 (first entry)

XX

DE Bovine EST associated with lactation/muscle/fat deposition #8781.

XX

KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
muscle deposition; fat deposition; genome mapping; gene identification;
gene analysis; cattle breeding.

XX

OS Bos Taurus.

XX

PN US2002137139-A1.

XX

PD 26-SEP-2002.

XX

PF 24-SEP-2001; 2001US-0960352.

XX

PR 12-JAN-1999; 99US-115707P.

XX

PR 11-JAN-2000; 2000US-0480902.

XX

(BYAT/) BYATT J C.

PA

(MATH/) MATHIALAGAN N.

PA

(TAON/) TAO N.

PA

(WARR/) WARREN W C.

XX

PI Byatt JC, Mathialagan N, Tao N, Warren WC;

XX

WPI; 2003-110599/10.

XX

DR New nucleic acid associated with lactation, and muscle and fat

XX

PT deposition, useful for genome mapping, gene identification and

PT

PT analysis, cattle breeding, or for genetically improving cattle

XX

PS Claim 2; SEQ ID No 8781; 245pp; English.

XX

CC The invention relates to a purified nucleic acid molecule associated with
lactation or muscle and fat deposition (designated LMFD), derived
from cattle, and the LMFD nucleic acid can specifically hybridise to a
second nucleic acid molecule comprising any of 15112 nucleotide
sequences, appearing as ABX4836-ABX4947, or complements of them.
Also included are; (1) a transformed cell having a nucleic acid
comprising an LMFD nucleic acid linked to a promoter and a 3' non-
translated sequence that functions in the cell to cause termination of
transcription and addition of polyadenylated ribonucleotides to a 3' end
of the mRNA molecule; and (2) determining a level or pattern of a
molecule in a bovine cell or tissue comprising: (a) incubating a marker
nucleic acid (comprising any of the 15112 nucleic acid sequences or its
complement or fragment) with a complementary nucleic acid molecule
obtained from the bovine cell or tissue, where hybridisation between the
marker nucleic acid and the complementary nucleic acid permits the
detection of the molecule; and (b) detecting the level or pattern of the
complementary nucleic acid, where the detection of the complementary
nucleic acid is predictive of the level or pattern of the molecule.
The LMFD nucleic acid is used for determining a level or pattern
of a molecule in a bovine cell or tissue. It is useful for genome
mapping, gene identification and analysis, cattle breeding, preparation
of constructs for use in cattle gene expression, or for genetically
improving cattle. The present sequence is one of the 15112 bovine
LMFD EST (expressed sequence tag) nucleic acids.

CC Note: The present sequence was not shown in the specification but
was obtained in electronic format from the USFTO web site:
seqdata.uspto.gov/sequence.html?DocID=20020137139.

XX

SQ Sequence 416 BP; 140 A; 77 C; 101 G; 98 T; 0 other;

XX

Query Match 12.1%; Score 123.2; DB 25; Length 416;

Best Local Similarity 78.2%; Pred. No. 5.4e-28;

Matches 186; Conservative 0; Mismatches 48; Indels 4; Gaps 3;

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 14:02:08 ; Search time 448.662 Seconds
(without alignments)
9221.962 Million cell updates/sec

Title: US-09-717-321A-2

Perfect score: 1266

Sequence: 1 ttttttttttttttttcaaa.....gagcaagaatagggtttt 1266

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	458.8	36.2	2941	14	US-10-198-846-9874
C 2	447.6	35.4	2051	9	US-09-925-302-255
C 3	430.6	34.0	1644	14	US-10-153-668-48
C 4	430.6	34.0	1646	14	US-10-153-668-46
C 5	420.2	33.2	1232	10	US-09-880-107-1632
C 6	284.2	22.4	2313	14	US-10-198-846-12919
C 7	195.2	15.4	594	14	US-10-066-543-2864
C 8	193.2	15.3	602	10	US-09-878-178-1654
C 9	193.2	15.3	602	13	US-10-046-935-1654
C 10	193.2	15.3	602	14	US-10-146-502-1654
C 11	187.8	14.8	549	14	US-10-066-543-2878
C 12	187.8	14.8	567	14	US-10-066-543-2864
C 13	187.2	14.8	605	11	US-09-871-161-483
C 14	185.4	14.6	477	14	US-10-066-543-3022
C 15	172.2	13.6	432	10	US-09-960-352-7612
C 16	164	13.0	2167	12	US-10-101-510-500

17	163.6	12.9	353	10	US-09-960-352-11432
C 18	147.4	11.6	409	14	US-10-066-543-13
C 19	126.4	10.0	285	9	US-09-294-093B-506
C 20	123.2	9.7	416	10	US-09-960-352-8781
C 21	116.4	9.2	3740	11	US-09-764-891-9984
C 22	108.8	8.6	293	14	US-10-066-543-3077
C 23	98.2	7.8	581	10	US-09-998-598-972
C 24	95.4	7.5	201	10	US-09-960-352-4357
C 25	91.6	7.2	565	10	US-09-998-598-296
C 26	85.6	6.8	492	10	US-09-878-178-1503
C 27	85.6	6.8	492	13	US-10-046-935-1503
C 28	85.6	6.8	492	14	US-10-146-502-1503
C 29	82.2	6.5	398	11	US-09-918-995-34569
C 30	68	5.4	588	11	US-09-871-161-133
C 31	62.2	4.9	426	10	US-09-960-352-7342
C 32	60	4.7	404	10	US-09-960-352-14206
C 33	57.6	4.5	1651	14	US-10-198-846-13019
C 34	52.4	4.1	162	10	US-09-920-300A-1349
C 35	52.4	4.1	162	12	US-10-099-926-1549
C 36	52.4	4.1	162	13	US-10-033-528-1549
C 37	52.2	4.1	424	10	US-09-960-352-11218
C 38	49.4	3.9	469	11	US-09-918-995-13017
C 39	48.6	3.8	336	12	US-09-814-353-4742
C 40	48.6	3.8	336	12	US-09-814-353-11039
C 41	48.2	3.8	277	10	US-09-960-352-12673
C 42	48	3.8	640681	10	US-09-790-988-1
C 43	47.8	3.8	2218	12	US-09-933-767-232
C 44	47.8	3.8	2218	14	US-10-023-282-103
C 45	46.8	3.7	302	12	US-09-814-353-5335

ALIGNMENTS

RESULT 1

US-10-198-846-9874/c
; Sequence 9874, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9874
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-9874

Query Match 36.2%; Score 458.8; DB 14; Length 2941;
Best Local Similarity 68.0%; Pred. No. 1.4e-113;
Matches 895; Conservative 0; Mismatches 337; Indels 84; Gaps 15;

QY	1	TTTTTTTTTTTTTTTCAAGTTCGAAAGACATTTTTTTTTTTTTTTTATGATTCAGG 60
DB	2481	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
QY	61	ATTATTAGTCATACATGCAAAACATCTGTAAGTCATTAGCAAAAGATCAATGTAA 120
DB	2421	ATTATTAGTCATACATGCAAAACATATGTAATGTCATTAGCAAAAGATCAATGTAA 2362
QY	121	AAACTCCACAATTCGCAACTGCAATGAAAAAGTTTCTTAGTGGTCGAAGGC 180

2361	AAACA	CTCCACA	AAATCTGC	AAATTTAAAA	AAATTTCTGTTAGTGGCTGAAGGGT	2302
181	CCAA	CACTGTGTTCTTTGCCAGT	GAGTTAGT	TTGTACAGAACGGCGTTAGCACTACGCGC	-- 238	
2301	CCCACG	CTGTATTCTCGCCAGT	GAGTTAAGTTTGTACAGAACATCGTCAGCACTAGCA	CAG 2242		
239	TTGAC	AGAACCTCACAGACCCAAAG	-----TACCGGAAGCATG	277		
2241	TTTAC	AGAACCTCACAGACCCAAAGGAACATCAATAGGCAAAAGCGACTACAGAGCGGTG	2182			
278	TGTCGG	CGTGGGTGAGTCTAGAGGGGGGGCATCAATCACATGACAGTGTGTGTA	CTCT 337			
2181	TGTCGG	CGTGGCGAGGTAAAGA	-GGGTCA	GTATTGGTCAAGTGACAGTGTGCGTAA	TCT 2123	
338	GGCA	AGACAGTGATG	-----TTTCAGAA	TATCTAAATAGTTTAAAA	380	
2122	GGCA	AGACAGTGATGTTAAGAAGGTTCATAGTTTAAGAA	TATCTAAATATTTTAAAA	2063		
381	CTGT	TAAGCCGACAGCTGATTTCTACACCCAGTTACTAGAAAAAGGAAGGAA	CACATTA 440			
2062	CTAT	AAAGCTGCAACACATGATTTTACACTAGTTACTAGAAAACTAAGGAAG	CACATT 2003			
441	GTCA	GTGAGTGA	GTAAAGGAAGTGA	AAAAAGCAAGCACTTCTACTATCTACAAAAAATC	500	
2002	ATTAG	CTCTGAATA	----AAGTAA	CATGGAAGCACTTTACTAATTCGACAAAAAACC	1948	
501	TCCGA	ATGCATATACAGAAAGATCTTATAGTACAGTTCAGACATATGCTCGTTAAGAA	G 560			
1947	TTCT	TAATGCATATTCAGAAAGATTTTATAACAAG	-GAGGCATATGCTCAGTCAGAAG	1889		
561	GGGT	CTCCTAAAGAAAGCACTTGTAAAGTTAGCAACTGTGAGGATGCCAGTTTAAATAT	620			
1888	GGGT	CTCTAAGAAAGCACTTACTAAGTTAGCGACTAACAGAAACACCAAGTTTAAAGAT	1829			
621	GGA	CTCAACGCCCATCTGGGAGGGHACAGAGGGGAAGGGGGCTCAAGAGACACT	680			
1828	GAAT	TAATGCCC	AATTTGGGGAGGCATGCGAGGTGTANGAGAAAGAAAGCTTAAGAA	1769		
681	GAT	AGATCGCCCATTTGTCTATCTACTGTTTGACAGAAATTAACCGTTAAAAAGCTTTAC	740			
1768	AACAT	TTTCTGATTAATCAACACTTTCTTTCATCATCTACTGCAATGACAGAAATTAAC	1709			
741	CCGT	GACACTTTTATTCAGTTGAA	---TTACTCCATGTACAATGTAGTGTAAATTAATCT	797		
1708	CTTT	TAGAGTTTTTACC	CGTCACACTTTTCATTCCTTGTACAATGTAGTGTGTA	---AATCT 1653		
798	CTACT	TCATATTA	-GTCAAAATAC	TGCTGTCTCCTTTGATGAGTGTGTTTACACAC	856	
1652	GCAC	TCGTATTTGTCAAAATAC	TGCTTTGTCTTTGATCA	-----CACACAC	1603	
857	TCCAC	CCAGCACCCAGACTAGGAA	CAGAAATACTTCGTTAGAGGCAACACAGGAGCCA	916		
1602	CCAC	CCGGCAACCCACAGCTA	--AACAGAA	TCTTCATTAGAGGAATA	-----GCA 1551	
917	GAGT	TCGTTCAAGCCTGCAAGCCGGTCA	GCTGGTATTTTAGAACTCACTATGAA	976		
1550	GTTC	TGTTTCAAAATCTCCGCAAAAGCTGGT	CAGAAAACTCGCTATGAAATCA	CAGAACT 1491		
977	ATCA	AGAGCAGAGCTGTTACACCCATCGTCACGTACAGTACAAAGTTA	CGTAAATGAGCA	1036		
1490	GAT	CAAGAGCTGAGCTGCTACGCTCACTCCATACAGTACANTGTTATGTCGGGAACA	1431			
1037	TGGG	CTGTATAGTTTACAGGTGGTTTACATGGCAGCGGTCA	TATTAAGAGAGGCTGTGCTGTG	1096		
1430	CGTG	CTGTAACTCACTGGTGAGTTCAAT	TGGCAACGCTTCATTCGGGAGGCTGTCTGCT	1371		
1097	TCAC	ACGGTCTGG	-----GAGCTACGGGAGGCTGCACC	-CCTGAGCCCAAGCTG 1148		
1370	TTAC	GCATCTGAGAACTCATAGAGAGCAAGTGTCTGCAC	CTCTTAATCTCAGAGCTA	1311		
1149	CAGT	CTCTTAAAGGACAAAGTCTCTCAACAGCTTAGTGC	TTTACGTGTTCTCAGCAACAC	- 1207		
1310	CGGT	CTCTCAAGAAGCAAGGCTTTTGGAAAGTTTCAGTGC	TTTCGGTGTCTCGGCAACCA	1251		

Query 1208 --GCAACTTAGTTCACAAAGTATTTGGCAATTCTTAATCTCGACAAGAATAGGGG 1261
||| ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dbb 1250 ATGCAGTGATGTTCAGGAAGTATTTTGGCAACTCTTAATCTGAACAAGAATGGGG 1195

RESULT 2

US-09-925-302-255/c ; Sequence 255, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIORITY FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-255

Query Match 35.4%; Score 447.6; DB 9; Length 2051;
Best Local Similarity 68.2%; Pred.No. 1.2e-110;
Matches 876; Conservative 0; Mismatches 324; Indels 84; Gaps 15;

Qy 33 TTTTTTTTTTTTTTTATGATTTCAAAGATTTATTAAGTCATCATGCCAAAAACATACTGC 92
Db 2009 TTTTTTTTTTTTTTTATGATTTCAAAGATTTATTAAGTCATCATGCCAAAAACATAATGC 1950

Qy 93 TAAC TGCAATAGCAAAAAGATCAATGTAAAACACTCCACAAATTCGCAACTGTCAATTGA 152
Db 1949 TAA TTGCAATAGCAAAAAGATCAATGTAAAACACTCCACAAATTCGCAACTGTCAATTGA 1890

Qy 153 AAAAGATTGTCTAGTGGTCGAAGGCCAACACTGTGTTCTTCGCAGTAGTTAGTTG 212
Db 1889 AAAAAATCTGTGTAGTGGTCGAAGGGTCCCACGCTGTA TTCTCGCAGTAGTTAGTT 1810

Qy 213 GTACAGAAGCGGTGTAGCACTAGCGC--TTGACAGAACCCTCACAGACCCAAAGG----- 264
Db 1829 GTACAGAATCTGTAGTGGTCGAAGGGTCCCACGCTGTA TTCTCGCAGTAGTTAGTT 1810

Qy 265 -----TACCGAAGCATGTGTCCGCGTGGGTGAGGTCTAGAGGGGGCGGC 309
Db 1769 AATAGGCAAGCGCACTACAGGAGCGGTGTGTCGCGTGGCGAGGTAAAGA-GGGTCAGT 1711

Qy 310 ATTCAATCATGACAGCTGTTCGTACTCTGGCAAGACAGTGAATG----- 352
Db 1710 ATTGTCAAGTACAGCTGTCCGTAATCTGGCAAGACAGTGAATGTAAGAAGTTCAVAGT 1651

Qy 353 TTTTCAGATATCTAAATAGTTTAAAAAAGTGTAAAGCGCAGCAGTGAATTTCTACACC 412

Db 1650 TTAAGAATTACTTAAATATTTTAAATACTATAAGCTGCAACATCATGTTTACACT 1591
QY 413 AGTTACTAGAAAACGAAGGAGACACTAGTCACTGAGTAAGAGAGGTGAAACAGAA 472
Db 1590 AGTTACTAGAAAACGAAGGAGACACTTATTAGTCTGAATA-----AAGTAACATGAA 1536
QY 473 CGCACTTCTACTATCTACCAAAAAATCTCCGATGCAATTATCAGAAGATCTTATAGTA 532
Db 1535 AGCACTTTTACTATCGCAAAAAACCTCTTAATGCAATTATCAGAAGATTTTATAATA 1476
QY 533 CAGTCTGACATATGCTCGTTAAAGGGGCTCTTAAGAAAAAGCACTTCTTAAGTTAG 592
Db 1475 CAAG-GAGGCATATTGCTCAGTCAGAAGGGTCTCTAATAAGAAAGCACTTACTAAGTTAG 1417
QY 593 CAATCTGTGAGATCGGCAGTTTAATATATGACTCAAGCCCTCTGCGGAGGACAGCA 652
Db 1416 CGACTAAGACAAACACAGTTTAAAGATGAATTAATGCCAATTTGGGAGGCATGCA 1357
QY 653 GGGGAAGGGGGCTCAAGAGACACTGATAAGATCGGCATTTGTCATCTACTGTTG 712
Db 1356 GGTTAGAGAAAGGAAAGCTTAAGAAACATTTCTGTATATACCAACCTTTCTTTCA 1297
QY 713 ACAGAAATTAACGTTAAAGCTTTTACCCTGTGACACTTTTATTCACTGAA---TTACT 769
Db 1296 TCATCTACTGCATTTGACAGAAATTAACCTTTTATAGAGTTTATCCCGTACACTTTCA 1237
QY 770 CCAATGACATATGCTGTAATTAATTAATCTTACTTCTATA-TTAGTCAAAATCTGCTGTC 828
Db 1236 CTTGTACAATGAGTGTGTA-----AATCTCCACTTCGTATTGTCAAAATCTGCTTTG 1181
QY 829 TCCCTTTGATGACCTCGTGTTCACACTCCACCAGCACACCACACTAGGACAGAA 888
Db 1180 TCCCTTGATCA-----CACACACCACCCCGCACACCACAGCTA--AACAGAA 1133
QY 889 TACTCTGTAGAGCAACACAGGAGCAGAGTTCTGTTCAAAGCCTGCAGAACCGGTCA 948
Db 1132 TTCTTCATTAGAGGAATA-----GCAGTTCTGTTCAAAATCTCCGCAAGCTGTCA 1079
QY 949 GCTGGTATTTTAGAGACTCACTATGAAATCAAGAGCAGAGCTGTTTACACCCATGTA 1008
Db 1078 GAAACTCGGTATGAAATCAAAAGACTGATCAAAAGAGCTGACTGCTCACTCC 1019
QY 1009 CGTACAGTCAAAAGTTACGTAATGAGCATGGGTGATAGCTTACAGTGGTTTACATGSC 1068
Db 1018 ATTACAGTCAATGTTATGTCGGGAACACGTGCTGCTACTCACTGCTGAGTTCAATGGC 959
QY 1069 AGCGTGTCAATTAAGAGGCTGCTGTGTACACGCTTGG-----GAGCTACGGGAG 1121
Db 958 AAGCCTTCATTCGGGAGGCTGTCTGCTTTACGCACTGAGAACTACATAGGAGCAAG 899
QY 1122 GGTCTGCACC-CCTGAGCCGAGAGCTGAGTCTTCTTAAGGCAAGTCTCTCAACAGC 1180
Db 898 TGTCTGCACCTCTAACTCGAGAAGTACCGTCTCTCAAGAGCAAGGCTCTTTGCAAG 839
QY 1181 TTAGTGTCTTACGTTTCTCAGACAAAC---GCAACTTGTTCACAGGTTATTTGCAAT 1237
Db 838 TTCAGTGTCTGGTGTCTCGGACAAACATGCAATGCAATGCTAGTTTCAAGAGGTTATTTGCAAC 779
QY 1238 TCTTAATCTGAGCAAGATAGGGG 1261
Db 778 TCTTAATCTGAACAGATGGGG 755

RESULT 3

US-10-153-668-48/c

; Sequence 48, Application US/10153668

; Publication No. US20030092616A1

; GENERAL INFORMATION:

; APPLICANT: HONDA, Goichi

; APPLICANT: MATSUDA, Akio

; APPLICANT: MURAMATSU, Shuji

; APPLICANT: ISHIZAWA, Kenya

; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
US-10-153-668-48

Query Match 34.0%; Score 430.6; DB 14; Length 1644;
Best Local Similarity 67.8%; Pred. No. 4.4e-106;
Matches 859; Conservative 0; Mismatches 354; Indels 84; Gaps 15;
QY 50 ATGATCAAGGATTTATTAAAGTCATACATGCAAAACATCTGCTAACTGCTAGCAAAA 109
Db 1644 ATGATCAAGGATTTATTAAAGTCATACATGCAAAACATATGCTAACTGCTAGCAAAA 1595
QY 110 GATCAATGTAAACACTCCACAATTCGCACTGCACTGCAATGAAAGTTTGTCTAGT 169
Db 1584 GATCAATGTAAACACTCCACAATTCGCACTGCACTGCAATGAAAGTTTGTCTAGT 1525
QY 170 GGTGGAAGGCCCAACACTGTGTTCTTCCAGTGAGTTAGTTGTACAGAACGGCTAG 229
Db 1524 GGTGGAAGGTCCTCAGCTGTATCTCCAGTGAGTTAGTTGTACAGAACATCGTCAG 1465
QY 230 CACTAGCGC--TTGACAGAACCTTCACAGACCCCAAGG-----TA 266
Db 1464 CACTAGCAGAGTTTACAGAACCTTCACAGACCCCAAGGCAACATCAATAGGCAAGCGACTA 1405
QY 267 CCGGAAGCATGTCTCCGCTGGGTGAGTCTAGAGGGGGGGGCAATCAATCAGATGACGT 326
Db 1404 CAGGAGGCTGTGTCCGCTGGGCGAGGTAAGA-GGGTCAGTATTGGTCAAGTGACAGT 1346
QY 327 GTTGTACTCTGCAAGACAGATGATG-----TTTCAGAATATCTTAAA 369
Db 1345 GTCCGTAACTCGCAAGACAGATGATGTTAAGAGGTTCTAGTTTAAGATTTATCTAAA 1286
QY 370 TAGTTTAAAACTGTAAAGCCGACAGCAGTGTCTACACCCAGTTTACTAGAAAAACGAA 429
Db 1285 TATTTTAAAACTATAAAGCTGCAACACATGATTTTACACCTAGTTTACTAGAAAACTAA 1226
QY 430 GGAAGCACTAGTCACTGAGTAAAGGAGGTGAAACAGGAACGCACTTCTACTATCTA 489
Db 1225 GGAAGCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAATCG 1171
QY 490 CCAAAAAAATCTCCGAATGCATTATCAGAAGATCTTATAGTACAGGTGACAGATATTGCG 549
Db 1170 ACAAAAAACCCTTCTAATGCAATTATCAGAAAGATTTTATATACAG-GAGGCATATTGCG 1112
QY 550 TCGTTAAGAGGGGTCTCTTAAAGAAAGCACTTGTCTAGTTAGCAACTCTGAGGATGGCC 609
Db 1111 TCAGTCAGAAGGGGTCTTATAGAAAGAAAGCACTTCTAAGTTAGGCACTACACAGAACACC 1052
QY 610 AGTTTAAATATGACTCAACGCCCATCTGGGAGGGGACAGCAGGGGGAAGGGGGCTCA 669

Db	1051	AGTTTAAAGATGAATTAATTCGCCAATTTGGGAGGCGATGCGAGGTGTAAAGAAAGGAA	992
Qy	670	AGAGAGACACTGATAAGATCGGCATTTGTCTACTCTGTTTGAAGAAATTAACCGTTA	729
Db	991	AAGCTTAAAGAAACATTTCTGATATACCAACCTTTCTTTCATCATCTACTGCAATTGA	932
Qy	730	AAAAGCTTTACCGTGACACTTTTATTCAGTTGAA---TTACTCCATGATCAATGTAGTG	786
Db	931	CAGAAATTAACCTTTTAGAGTTTTTACCCGTGACACTTTTCATTCCTGTCAATGTAGTG	872
Qy	787	TAAATTAATCTCTACTTTCATA-TTAGTCAAAATACTGTCTGTCTCTTTTGATGACGTGT	845
Db	871	TA-----AATCTCCACTCTGTATTTTGTCAAAATACTGTCTTGTCTCTTTGATCA	822
Qy	846	GTTCACACACTCCACCCAGCACACCCAGCTAGGAAACAGAAATCTTCGTTAGAGGAA	905
Db	821	-----CACACACCCACCCGGCACACCCACAGCTA--AACAGAAATCTTCATTAGAGGAA	768
Qy	906	CACAGGACCGAGATTTCTGTTCAAGCCTGCAAGAGCCGGTCAGCTGGTATTTTAGAGAA	965
Db	767	TAG-----CAGTTCTGTTCAAAATCTCCGCAAAAGCTGGTCAGAAATCTGCTATGAAA	714
Qy	966	CTCACTATGAATCAAGAGCAGAGCTGTTACACCCATCGTGACGTACAGTACAAAGTTA	1025
Db	713	TCACAAAGACTGATCCAAAGAGCTGAGCTGTACGCTACTCCATTCAGTACATGTTA	654
Qy	1026	CGTATGAGCATGGGCTGATTAAGTTTACAGGTGCGTTTACATGCGACGGTGTCTAATAGGAG	1085
Db	653	TGTCGGAAACAGCTGCTCTACTCAGTGTGAGTTCATGCAACGCTTCATTCCGGGAG	594
Qy	1086	GCTGTGCTGTGTACACGGTCTGG-----CAGTACGGGAGGTCTGCACC-CCTGAG	1137
Db	593	GCTGTCTGCTTTAGCGATCTGAGAACTACATAGGAGCAAGTGTCTGCACCTCCTAAC	534
Qy	1138	CCGAGAGCTGAGTCTTCTTAAGACAAAGTCTCTACAGCTTAGTCTTAAGTGTTC	1197
Db	533	TCGAGAGCTACCGTCTTCTCAAGACGAAAGTCTTTGCAAGTTCAGTGTCTGCTGTTTC	474
Qy	1198	TCAGCACACAC---GCAACTTAGTTTCAAGAGTATTTTGGCAATTTCTTAATCTGAGCAAGA	1254
Db	473	TCGGCACACAAATGCACTGATGTTTCAAGAGGTATTTTGGCAACTCTTAAATCTGAACAAGA	414
Qy	1255	ATAGGGG 1261	
Db	413	ATGGGGG 407	
RESULT 4			
US-10-153-668-46/c			
; Sequence 46, Application US/10153668			
; Publication No. US2003092616A1			
; GENERAL INFORMATION:			
; APPLICANT: HONDA, Goichi			
; APPLICANT: MATSUDA, Akiyo			
; APPLICANT: MURAMATSU, Shuji			
; TITLE OF INVENTION: STAT6 Activating Gene			
; FILE REFERENCE: 1254-0207P			
; CURRENT APPLICATION NUMBER: US/10/153,668			
; CURRENT FILING DATE: 2002-05-24			
; PRIOR APPLICATION NUMBER: US 60/293,172			
; PRIOR FILING DATE: 2001-05-25			
; PRIOR APPLICATION NUMBER: US 60/316,031			
; PRIOR FILING DATE: 2001-08-31			
; PRIOR APPLICATION NUMBER: US 60/328,403			
; PRIOR FILING DATE: 2001-10-12			
; PRIOR APPLICATION NUMBER: JP 2001-157043			
; PRIOR FILING DATE: 2001-05-25			
; PRIOR APPLICATION NUMBER: JP 2001-260681			
; PRIOR FILING DATE: 2001-08-30			
; PRIOR APPLICATION NUMBER: JP 2001-313175			
; PRIOR FILING DATE: 2001-10-10			
; NUMBER OF SEQ ID NOS: 488			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 46			
; LENGTH: 1646			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (109)..(273)			
US-10-153-668-46			
Query Match			
Best Local Similarity 34.0%; Score 430.6; DB 14; Length 1646;			
Matches 859; Conservative 67.8%; Pred. No. 4.4e-106;			
Mismatches 324; Indels 84; Gaps 15;			
Qy	50	ATGATTCAGGATTTATTAAGTCAATGCAATCTCCCAATTTCTGCAACTCTCAATTTGAAAGAAAGTTTGTCTAGT	109
Db	1646	ATGATTCAGGATTTATTAAGTCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT	1587
Qy	110	GATCAATGTAAGAAACACTCCCAATTTCTGCAACTCTCAATTTGAAAGAAAGTTTGTCTAGT	169
Db	1586	GATCAATGTAAGAAACACTCCCAATTTCTGCAACTCTCAATTTGAAAGAAAGTTTGTCTAGT	1527
Qy	170	GGTCCAAAGGCCCAACACATGTGTTCTTGGCAGTGAAGTTAGTTGTACAGAACGGGTTAG	229
Db	1526	GGCTGAAGGGTCCCAACGCTGATTTCTCGCAGTGAAGTTAGTTGTACAGAACCATCGTCAG	1467
Qy	230	CATAGCGC--TTGACAGAACTCCACAGACCCCAAGG-----TA	266
Db	1466	CATAGCAGAGTTTACAGAACTCCACAGACCCCAAGGAAACATCAATAGGCAAGGCACTA	1407
Qy	267	CCGGAAGCATGTGTCGGGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCACATGACACT	326
Db	1406	CAGGAGGCGTGTGTCGGGTGGCGAGGTAAGA--GGGTCAATTTGGTCAAGTGACACT	1348
Qy	327	GTGGTACTCTGGCAAGACAGTCAAGT-----TTTCAGAAATATCTAAAA	369
Db	1347	GTGGTAACTCTGGCAAGACAGTCAAGTAAAGAGGTTTCAAGTAAAGAAATATCTAAAA	1288
Qy	370	TAGTTTAAAGACTGTAAGACCCGACGAGTATTTCTACACCCAGTACTAGAAAAAGAA	429
Db	1287	TATTTTAAAGACTGTAAGAGCTGCAACACATGATTTTACCTAGTACTAGAAAACTAA	1228
Qy	430	GGGAAGCACTAGTCAAGTGAAGTAAAGAGAGGTGAAAGAGGCAAGCGCACTTCTACTACTA	489
Db	1227	GGAAAGCACTTATAGTCTGAATA-----AAGTAACATGTAAGAGCACTTTACTAATCG	1173
Qy	490	CCAAAAAATCTCCGATGCAATTAAGAAAGATCTTATAGTACAGGTGACAGATATGC	549
Db	1172	ACAAAAAACCCTTCTAATGCATTTACAGAAAGATTTTATATAACAAG--GAGGCATATTC	1114
Qy	550	TCGTTAAGAGGGGTCTTAAAGAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGGCC	609
Db	1113	TCAGTCAGAAAGGGTTCCTAAGAAAGCACTTACTAAGTTAGCGACTACAGAAACAAC	1054
Qy	610	AGTTTAAATATGGACTCAACCGCCCATCTGGGAGGAGCAGAGGGGGAAGGGGGCTCA	669
Db	1053	AGTTTAAAGATGAATTAATGCCCCTTTGGGAGGAGTGGGAGGAGTGAAGAGAAAGGAA	994
Qy	670	AGAGAGACACTGATAAGATCGGCCATTTGTCTACTGTGTTTACAGAGAAATTAACCGTTA	729
Db	993	AAGCTTTAAGAAAAACATTTCTGTATAACCAACCTTTCTTCTATCATCTCTCATTTGA	934
Qy	730	AAAAGCTTTACCGTGCACACTTTTATTCAGTTTCAA---TTACTCCATGTACAATGATG	786
Db	933	CAGAAATTAACCTTTTAGAGTTTATACCGGTGCACACTTTCATTTCTTGTACATGATG	874
Qy	787	TAAATTTAATCTCTACTTCTATA-TTAGTCAAAATACTGTCTGTCTCTCTTGTATGACGTCGT	845
Db	873	TA-----AATCTCCACTTCGTATTTTGTCAAAATACTGTCTTGTCTCTTTGATCA	824
Qy	846	GTTCACACTCTCCACCCAGCACACCCAGCTAGGAAACAGAAATACTCTGTTAGAGCAA	905
Db	823	-----CACACACCCACCCGGCACACCCACAGCTA--AACAGAAATTTCTTATTAGAGGAA	770

QY 906 CACAGAGCCAGAGTTCTGTTCAAAGCCTGCAGAGCCGGTCAGCTGGTATTTTAGAGAA 965
Db 769 TAG-----CAGTTCTGTTCAAATCTCCGAAAAGCTGGTCAGAAAACCTCCCTATGAA 716
QY 966 CTCATATGAATCAAAGACGAGAGCTGTTTACACCCATCGTCAGCTACAGTACAAAGTTA 1025
Db 715 TCACAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTCCATTACAGTACAAATGTTA 656
QY 1026 CGTAATGAGCATGGCTGATAGTTACAGTTCGGTTCAATGGCAGCGTGTCTATTAAAGGAG 1085
Db 655 TGTGGGAACACGCTGCTGCTAACTCACTGCTGAGTTCATGCAACGCTTCATTCCGGAG 596
QY 1086 GCTGTGCTGTGTCACACGCTCTGG-----GAGCTACGGGAGGCTGTCACC-CCTGAG 1137
Db 595 GCTGTCTGCTTTAGCATCTGAGAACTACATAGAGAGCAAGTGTCTGCACTCTTAAC 536
QY 1138 CCGAAGCTGAGTCTTCTTAAGCAAAAGTCTCTCAACAGCTTAGTGCTTACGTGTTTC 1197
Db 535 TGCAGAAGCTACCGTCTTCTCAAGACGAAGTCTTTGCAAAAGTTTCAGTGTCTCGTGTTC 476
QY 1198 TCAGACAAAC---GCAACTTAGTTACAAAGGTTATTTGGCAATCTTTAATCTGAGCAAGA 1254
Db 475 TCGGACAAACAAATGCAAGTGTAGTTCAAGAGGTTATTTGGCAACTCTTAATCTGAACAAGA 416
QY 1255 ATAGGGG 1261
Db 415 ATGGGGG 409

RESULT 5

US-09-880-107-1632/c
; Sequence 1632, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1632
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D25274
US-09-880-107-1632

Query Match 33.2%; Score 420.2; DB 10; Length 1232;
Best Local Similarity 67.5%; Pred. No. 2.5e-103;
Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;
QY 53 ATTCAAGAGTTTATTAAAGTCATACATGCAAAACATACCTGCTAACTCATCTAGCAAAAGAT 112
Db 1228 WTCCAGGATTTATTAAAGTCATACATGCAAAACATATGCTAAATTCATTTAGCAAAAGAT 1169
QY 113 CAATGTAAAAACACTCCCAATTTCTCAACTGTCATTTGAAAAAGTTGTTCTAGTGGT 172
Db 1168 CAATGTAAAAACACTCCCAATTTCTCAACTGTCATTTGAAAAAGTTGTTCTAGTGGC 1109
QY 173 CGAAGGCCCAACACTGTGTTCTTGGCAGTGTAGTTAGTTGTACAGAACGGCTTAGCAC 232
Db 1108 TGAAGGGTCCCAAGCTGTATTTCTCGCCAGTGTAGTTAGTTGTACAGAACATCGTCAGCAC 1049

QY 233 TAGCGG--TTGCAGAACTCTCAGACCCAAAGG-----TACCG 269
Db 1048 TAGCACAGTTTACAGAACTCTCAGACCCCAAAGGAACATCAATAGGCAAGCGACTACAG 989
QY 270 GAAGCATGTGTCGGTGGTGGTGGTCTAGAGGGGGCGGCATCAATCAATGACAGTGT 329
Db 988 GAGGCGTGTGTCGGTGGTGGGAGGTAAGA--GGGTCAGTATTGGTCAAGTACAGTGT 930
QY 330 GGTACTCTGGCAAGACAGTGTG-----TTTCAGAAATCTAAATAAG 372
Db 929 GGTATCTGGCAAGACAGTGTGTTAAGAGTTTCAATAGTTTAAAGATTTCTAAATAT 870
QY 373 TTTAAAACTGTAAAGCCGAGCAGTGTATTTCTACCCAGTTTACTAGAAACGAAAGG 432
Db 869 TTTAAAACTATAAGCTGCAACACATGATTTTACCTAGTTTACTAGAAACGAAAG 810
QY 433 AAGCACTAGTCAAGTGTGTAAGGAAGGTGAACAGAGAAAGCACTTCTACTATCTACA 492
Db 809 AAGCACTTATAGTCTGAATA-----AAGTAACATGGAAAGCACTTTTACTTAATCGACA 755
QY 493 AAAAATCTCCGAATGCATTTATCAGAAAGATCTTATAGTACAGGTCAGACATATTGCTCG 552
Db 754 AAAAATCTTCTAATGCATTTATCAGAAAGATTTTATAATACAAG--GAGGCAATTTGCTCA 696
QY 553 TTAAGAGGGGTCTTAAGAAAAAGCACTTCTAAGTTAGCAACTGTGAGGATGCCAGT 612
Db 695 GTCAGAGGGGTCTTATAAGAAAAAGCACTTACTAAGTTAGCGACTTAACAGAAACACCACT 636
QY 613 TTAATATGGACTCAAGCCCATCTGGGAGGAGACAGCAGGGGAGGGGGGTCAAGA 672
Db 635 TTAAGATGAATTAATATGCCCAATTTGGGAGGATGCGAGGTGTAGAGAGAGGAAAG 576
QY 673 GAGCACTGATAGATCGGCCATTTGTCATCTACTGTTTGACAGAAATTAACCGTTAAAA 732
Db 575 CTTAAGAAAAACATTTCTCTGATATACCAACCTTTCTTTCATCTCTCTGATTTGACAG 516
QY 733 AGCTTTACCGGTGACACTTTTATTCAGTTGAA-----TTACTCCATGTACAAATGAGTAA 789
Db 515 AAATTAACCTTTTAGAGTGTTCACCGGTGACACTTTTACCTTTGTACAAATGAGTGA- 457
QY 790 ATTAATCTCTACTTCATA--TAGTCAAAATACTGTCTGTCTCTTTGATGACGCTGTGT 848
Db 456 ---AATCCACTTCGTATTTTGTCAAAATACTGTCTTTGCTCTTTGATCA----- 409
QY 849 TCACACTCTCCACCCAGCACACCCAGCTAGGAACAGAAATACTTCTGTAGAGCAACAC 908
Db 408 -CACACACCCACCCGCGCACACCCACAGCTA--AACAGAAATCTTCAATTAGAGGAATAG 352
QY 909 AGGAGCCAGAGTCTGTTCAAAGCCTGCAAGCCGCTCAGCTGGTATTTTAGAGAACTC 968
Db 351 -----CAGTTCTGTTCAAATCTCCGCAAAAGCTGGTCAGAAAACCTCGCTATGAATCA 298
QY 969 ACTATGAATCAAGAGCAGAGCTGTTACACCCATCGTGCAGTACAGTACAAAGTTACGT 1028
Db 297 CAAGAAGTCAATCAAGAGCTGAGCTGCTACGCTCACTCCATTTACAGTACAAATGTTATGT 238
QY 1029 AATGAGCATGGCTGATAGTTTACAGTGGTGTACATGGCAGCGTGTCAATTAGGAGGCT 1088
Db 237 CGGGAACACGTGCTGCTAATCACTGCTGAGTTCAATGGCAACGCTTCATTCGGGAGGCT 178
QY 1089 GTGCTGTGTACACGGTCTGG-----GAGTTACGGGAGGCTCTGCACC--CCTGAGCCC 1140
Db 177 GTTCTGCTTTAGCGATCTGAGAACTACATAGGAGAGCAAGTGTCTGCACCTCTTAAGTGC 118
QY 1141 AGAAGCTGAGTCTTCTTAAAGACAAAGTCTCTCAACAGCTTAGTGTCTAGTGTCTCA 1200
Db 117 AGAAGCTACGCTTCTCTCAAGACGAAAGTCTTTGCTAAAGTTCAAGTGTCTCGTGTCTCG 58
QY 1201 GCACAAC---GCAACTTAGTTTCAAGAGTATTTTGGCAATCTTAACTCTGAGCAAGA 1254
Db 57 GCACAACATGCAAGTGTAGTTTCAAGAGGTTATTTGGCAACTCTTAACTCTGAACAAGA 1

RESULT 6

US-10-198-846-12919/c
; Sequence 12919, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12919
; LENGTH: 2313
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-12919

Query Match 22.4%; Score 284.2; DB 14; Length 2313;
Best Local Similarity 63.5%; Pred. No. 2.9e-66;
Matches 651; Conservative 0; Mismatches 318; Indels 57; Gaps 12;

QY	265	TACCGAGACAGTGTCCGCGTGGTGGTGTAGAGGGGGGGGCAATCAATCACATGACA	324
DB	2221	TACAGAGCGGTGTGCCCGTGGCGAGGTAAAG-GGGTCAGTATGGTCAAGTGACA	2163
QY	325	GTGTTGGTACTCGGCAACACAGTGTATTTTCAGAA-----TATCTAA	367
DB	2162	GTGTCGTAACTGGCAACACAGTGTATTTTCAGAAAGTTTCAAGTATTAAGTAA	2103
QY	368	ATAGTTTAAACTGTAAAGCGGACAGTGTATTTTCACCCAGTTACTAGAAACG	427
DB	2102	ATATTTTAAACTATAAGCTGCAACATGATTTTACACCTAGTTACTAGAAACT	2043
QY	428	AAGGAGACACTAGTACAGTGTAGTAAAGGAGTGAACAGGACGCTTCTCATATC	487
DB	2042	AGGAAAGCACTTATAGCTCTGAATA-----AAGTAACTGGAAGACCTTTACTAAT	1988
QY	488	TACCAAAAAATCTCCGAATGCAATTATCAAGAAAGATCTTATAGTACAGTFCAGACATAT	547
DB	1987	CGACAAAAAACCTTCTAATGCATTATCAGAAAGATTTTATAATAACAAG-GAGGCATAT	1929
QY	548	GCTCGTTAAGAGGGGGTCTTAAGAAAGCACTTCTAAGTTAGCAACTGTGAGGATGG	607
DB	1928	GCTCAGTCAAGAGGGGTCTATAAGAAAGCACTTACTAAGTTAGGCACTTAACAGAACAA	1869
QY	608	CCAGTTTAAATATGGACTCAACCCCATCTGGGAGGACAGCAGGGGGAAGGGGGCT	667
DB	1868	CCAGTTTAAAGATGAATTAATCCCAATTTGGGAGGCATGCCAGGTGTAAAGAAAG	1809
QY	668	CAAGAGACACTGATAAGATCGGCCATTTGTCTACTCTGTTTGTACAGAAATTAACCGT	727
DB	1808	AAAAGCTTAAAGAAACATTTCTCTGATTAATCAACCAACCTTTCTTCATCATCTACTGCAAT	1749
QY	728	TAAAAGCTTTACCGTGACACTTTTATTCAGTTGAATTAATCTCATCTGATCAATAGTAGT	787
DB	1748	GACAGAAATTAACCTTTTAGAGTTTTTACCCGTCACACT--TTCTTCCTTTGACAAATG	1691
QY	788	AAATTAATCTCTACTTCATTA--TTAGTCAAAATACTGTCTGTCTCTCTTTGATGACGTGGT	846
DB	1690	AGTGTAACTCCACTTCGTATTTTGTCAAAATACTGTCTTTTGTCTTTGATCA-----	1638
QY	847	TTTCACACTCCACCCAGACACCCACGACTAGGAACAGAAATCTTCGTTAGGGAAC	906
DB	1637	---CACACACCCACCCGGCACACCAAGCTA--AACAGAAATCTTCTATTAGAGGAAT	1583

QY	907	ACAGGAGCCAGAGTTCTGTTCAAAAGCCTGCAGAACCGGTGAGCTGGTATTTTAGAGAAC	966
DB	1582	A-----CAGTTCTGTTCAAAATCTCCGCAAAAGCTGGTCAGAAAACTCGTATGAAAT	1530
QY	967	TCACTATGAATCAAAAGACAGAGCTGTTTACCCCATCGTGACGTACAGTACAAAGTTAC	1026
DB	1529	CACAAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTCCATTCACAGTACAATGTTAT	1470
QY	1027	GTAATGACATGGCTGATAAGTTACAGTTCGTTTACATGGCAGCGTGTCTATTAAGGAGG	1086
DB	1469	GTCCGGAACACGTGCTGCTACTACTCTGTTGAGTTCAATGGCAACGCTTCATTCGGGAG	1410
QY	1087	CTGTCTGTGTACACACGCTCTGG-----GAGCTACGAGGAGGCTGTGCACC-CTGAGC	1138
DB	1409	CTGTTCTGCTTTACGCATCTGAGAACTACATAGGAGAGCAAGTGTCTGCACCTCTAACT	1350
QY	1139	CCAGAAAGCTGAGTCTTTTAAGGACAAAGTCTCTCAACAGCTTAGTGTAGCTGTTCT	1198
DB	1349	GCAGAAAGCTACCGTCTTCTCAAAAGACGAGGTCTTTGCAAAAGTTTCAAGTCTCGGTGTTCT	1290
QY	1199	CAGCACAAAC---GCAACTTAGTTCAAGGTATTTTGGCAATTTCTTAATCTGAGCAAGAA	1255
DB	1289	CGGCACAAACATGCAAGTGTAGTTCAAGGTATTTTGGCAACTTTTAATCTGACAAAGAA	1230
QY	1256	TAGGGG 1261	
DB	1229	TGGGG 1224	

RESULT 7

US-10-066-543-2864/c
; Sequence 2864, Application US/10066543
; Publication No. US2003008781A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2864
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2864

Query Match 15.4%; Score 195.2; DB 14; Length 594;
Best Local Similarity 70.9%; Pred. No. 1.9e-42;
Matches 373; Conservative 0; Mismatches 113; Indels 40; Gaps 7;

QY	265	TACCGAAGCATGTCCGCGTGGGTGAGGTCTAGAGGGGGGGGCAATCAATCATGACA	324
DB	524	TACAGAGGCGTGTGTCGCGTGGCGAGGTAAAGA-GGTCAGTATTTGCTCAAGTGACA	466
QY	325	GTGTGTACTCTGGCAAGACAGTGTATTTTCAGAA-----TATCTAA	367
DB	465	GTGTGCGTAACTCTGGCAAGACAGTGTATTTTAAGAGGTTTCATAGTTTAAAGATTTATCTAA	406
QY	368	AATAGTTTAAAACTGTAAAGCGGACAGCTGATTTCTACCCAGTTACTAGAAAACG	427
DB	405	AATATTTTAAAACTATAAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC	346

Qy	428	AAGGGAACCACTAGTCAGCTAGTAAAGGAGGTAAGAAACAGCAACGCACTTCTACTATC	487
Db	345	AAGGAACCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAAT	291
Qy	488	TACCAAAAAATCTCCGAATGCATTATATCAGAAGATCTTTATAGTACAGGTACAGACATATT	547
Db	290	CGACAAAAAAACCTTCTAATGCATTATCAGAAGAATTTTATAATCAAG-GAGGCATATT	232
Qy	548	GCTCGTTAAGAGGGGGTCTTAAGAAAAAGCACTTCTTAAGTTAGCACTGTGAGATGG	607
Db	231	GCTCAGTCAGAGGGGTCTTATAAGAAAGCACTTACTAAAGTTAGCGACTAACAAGACAA	172
Qy	608	CCAGTTTAAATATGAACTCAACGCCCTCTCGGAGGACAGCAGG-----GGG	657
Db	171	CCAGTTTAAAGATGAATTAATGCCCANTTTGGGAGGCATGCGAGGTGTAGAGAAAGG	112
Qy	658	AAGGGGGCTCAAGAGAGACACTGATAAGATCGGC-----CATTTGTCATCTACTG--TTT	711
Db	111	AAAAGCTTTAAGAAAAACATTTCTGTATAATACCAACCTTCTTTTCATCATCTACTGCATTT	52
Qy	712	GACAGAAATTAACCGTTAAAAAGCTTTTACCGTGACATTTTATTC	757
Db	51	GACAGAAATTAACCTTTTAGAGTTTTTTACCGGTGACATTTTCATTC	6

```

RESULT 8
US-09-878-178-1654
; Sequence 1654, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Xuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(602)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1654

```

	Query Match	15.3%;	Score 193.2;	DB 10;	Length 602;
	Best Local Similarity	70.5%;	Pred. No. 6.7e-42;		
	Matches 371;	Conservative 0;	Mismatches 115;	Indels 40;	Gaps 7;
QY	265	TACGGAGCATGTCCGGTGGTCAGGCTCTAGAGGGGGCGCATCAATCAGTACA	324		
Db	74	TACAGAGGCGTGTGTCCGGTGGCGAGGTAAAGA--GGTCACTATTGGTCAAGTGCCA	132		
QY	325	GTGTGGTACTCTGGCAAGACAGTGAATGTTTCAGAA-----TATCTAA	367		
Db	133	GTGTCCGTATCTGGCAAGACAGTGAATGTTTAAGAAGGTTCATAGTTTAAAGNAITTTATCTAA	192		
QY	368	AATAGTTTTAAAACTGTAAAGCCGACAGCTGATTTCTACCCAGTGTACTAGAAAACG	427		
Db	193	AATATTTTAAAACTATAAAGCTGCAACACATGATTTTACACCTAGTTTACTAGAAAAC	252		
QY	428	AAGGGAGCACTAGTCACTGTAGTAAAGGAAGGTGAAAACAGGAAACCACTTCTACTATC	487		
Db	253	AAGGAAAGCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTTAAT	307		
QY	488	TACCAAAAAAATCTCCGAATGCATTAATCAGAAAGATCTTATAGTACAGGTCAGACATATT	547		

```

308 CGACAAAAAACCTTCTTAATGCATTATCAGAAGATTTTTATAATACAAG-GAGGCATATT 366
548 GCTCGTTAAGAGGGGGTCTAAAGAAAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGG 607
367 GCTCAGTCAGAAGGGTTCCTATAGAAAAGCACTTACTTAAGTTAGCGACTAACAGAACAA 426
608 CCAGTTTAAATATGACTCAAGCCGCCCTCTGGGAGGAGACAGAGG-----GGG 657
427 CCAGTTTAAAGATGAATTAATGCCCAATTTGGGAGGCATGGCAGGTGTAAGAGAAAGG 486
658 AAGGGGGGCTCAAGAGAGACACTGTATAAGATCGGC-----CATTTGTCATCTACTG---TTT 711
487 AAAGCTTAAAGAAACATTTCTCTGATATACCAACTTTCITTCATCTACTGCAATTT 546
712 GACAGAAATTAACCGTTAAAAAGCTTTACCCGTGACACTTTTATTC 757
547 GACAGAAATTAACCTTTTAGAGNTTTTACCNTGACACTTTCATTTC 592

RESULT 9
US-10-046-935-1654
; Sequence 1654, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 569, 578
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-1654

```

Query Match	15.3%;	Score 193.2;	DB 13;	Length 602;		
Best Local Similarity	70.5%;	Pred. No. 6.7e-42;				
Matches 371; Conservative	0;	Mismatches 115;	Indels 40;	Gaps 7		
QY	265	TACCGNAGCATGTGTC	CGCGTGGGT	CAGGTC	TACAGAGGGGGCGGCATCAATCAGATGACA	324
Db	74	TACAGGAGGCGTGTG	TCGCGGTGGCGAGGTAAAGA	-GGGT	CAGTATTGGTCAAGTGCCA	132
QY	325	GTGTTGTCATCTCGG	CAAGACAGT	GTGTTTCAGAA	-----TATCTAA	367
Db	133	GTGTCGTAATCGGC	AAAGACAGT	GTGTTAAGNAGGTT	CATAGTTAAGNAATTATCTTAA	192
QY	368	AATAGTTTTAAAACT	GTAAAGCCGAC	ACGTGATTTCTACCCAGT	TCTACTAGAAAAACG	427
Db	193	AATATTTTTAAAACT	ATAAAGCTGCAAC	CATGATTTTTACCTAGT	TCTACTAGAAAACT	252
QY	428	AAGGGAAAGCACTAGT	CAGCTGAGTAAAG	GAAGGTGAAACAGGAAC	CGCATTTCTACTATC	487
Db	253	AAGAAAGCACTATT	TAGCTCTGAATA	-----AAGTAACATG	GAAGACATTTTACTAAT	307
QY	488	TACCAAAAAAATCT	CCGAATGCATTAT	TCAGAAAGATCTTTAT	GTACAGGTCAGACATATT	547
Db	308	CGACAAAAAAACCT	TCTTAATGCAATTAT	CAGAAAGATTTTAT	TATATACAG-GAGGCATATT	366
QY	548	GCTCGTTAAGAGGG	GGTCTTAAAGAAAGCA	CTTGCTAAGTTAG	CAACTGTGAGGATGG	607
Db	367	GCTCAGTCAGAGGG	GTCTTATAAGAAAGCA	CTTACTAAGTTAGG	CACTTAAAGAACAA	426

RESULT 12

US-10-066-543-2964
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066.543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2964

Query Match 14.8%; Score 187.8; DB 14; Length 567;
Best Local Similarity 73.3%; Pred. No. 1.9e-40;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

QY 265 TACCGNAGCATGTGTCGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 324
Db 74 TACAGGAGGGGTGTGTCGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 132

QY 325 GTGTGCTACTCTGGCAAGACAGTGATGTTTCAGAA-----TATCTAA 367
Db 133 GTGTGCTAATCTGGCAAGACAGTGATGTTTCAGAAAGGTTTATAGTTAAGAAATATCTAA 192

QY 368 AATAGTTTAAAACTGTAAAGCCGACGACGTGATTTCTACCCAGTTACTAGAAAACG 427
Db 193 AATATTTTAAAACTATAAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC 252

QY 428 AAGGAGGACACTAGTCAGCTGAGTAAAGGAGGTGAAACAGGAGCCACTTCTACTATC 487
Db 253 AAGGAAGGACCTTATTAGCTCTGAATA-----AAGTAACATGGAAGACACTTTTACTAAT 307

QY 488 TACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTACAGATATT 547
Db 308 CGACAAAAAACCTTCTTANTGCAITATCAGAAAGATTTTATAATACAG-GAGGCATATT 366

QY 548 GCTCGTTAAGAGGGGTCTTAAAGAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGG 607
Db 367 GCTCAGTCAGAGGGGTCTTATAAGAAAGCACTTACTAAGTTAGCGACTAACAGAACAA 426

QY 608 CCAGTTTAAATATGGACTCAACGCCCTCTGGGAGGGACAGCAGGGGGAAGGGGGCT 667
Db 427 CCGTTTAAAGATGAATTAATGCCCATTTGGGGANGCATGCCAGGTGTGTAAGANAAN 486

QY 668 CAAGAGAGACACTGATAAGATCGGCCATTGTG-ATCTACTGTTTCAGAGAAATTAACCG 726
Db 487 GAAAGCNTAAGAAACATTTNCTGGTTATANCAACCTTTNTTNTTATCTACTGNATT 546

QY 727 TTAAGAGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAC 777
Db 547 TGACAAAAATTAACCNCTTTAAAGTTTTTACCCNGGCACCTNNCTNTCTGTGCC 597

RESULT 13

US-09-871-161-483
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun

; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/871.161
; CURRENT FILING DATE: 2001-05-31
; PRIOR FILING DATE: 1999-06-08
; PRIOR FILING DATE: 1999-06-08
; PRIOR FILING DATE: 1999-01-27
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

Query Match 14.8%; Score 187.2; DB 11; Length 605;
Best Local Similarity 65.5%; Pred. No. 2.9e-40;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

QY 265 TACCGAAGCATGTGTCGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 324
Db 74 TACAGGAGGGGTGTGTCGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 132

QY 325 GTGTGCTACTCTGGCAAGACAGTGATGTTTCAGAA-----TATCTAA 367
Db 133 GTGTGCTAATCTGGCAAGACAGTGATGTTTCAGAAAGGTTTATAGTTAAGAAATATCTAA 192

QY 368 AATAGTTTAAAACTGTAAAGCCGACGACGTGATTTCTACCCAGTTACTAGAAAACG 427
Db 193 AATATTTTAAAACTATAAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC 252

QY 428 AAGGAGGACACTAGTCAGCTGAGTAAAGGAGGTGAAACAGGAGCCACTTCTACTATC 487
Db 253 AAGGAAGGACCTTATTAGCTCTGAATA-----AAGTAACATGGAAGACACTTTTACTAAT 307

QY 488 TACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTACAGATATT 547
Db 308 CGACAAAAAACCTTCTTANTGCAITATCAGAAAGATTTTATAATACAG-GAGGCATATT 366

QY 548 GCTCGTTAAGAGGGGTCTTAAAGAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGG 607
Db 367 GCTCAGTCAGAGGGGTCTTATAAGAAAGCACTTACTAAGTTAGCGACTAACAGAACAA 426

QY 608 CCAGTTTAAATATGGACTCAACGCCCTCTGGGAGGGACAGCAGGGGGAAGGGGGCT 667
Db 427 CCGTTTAAAGATGAATTAATGCCCATTTGGGGANGCATGCCAGGTGTGTAAGANAAN 486

QY 668 CAAGAGAGACACTGATAAGATCGGCCATTGTG-ATCTACTGTTTCAGAGAAATTAACCG 726
Db 487 GAAAGCNTAAGAAACATTTNCTGGTTATANCAACCTTTNTTNTTATCTACTGNATT 546

QY 727 TTAAGAGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAC 777
Db 547 TGACAAAAATTAACCNCTTTAAAGTTTTTACCCNGGCACCTNNCTNTCTGTGCC 597

RESULT 14

US-10-066-543-3022
; Sequence 3022, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun

; APPLICANT: Indrias, Carol Yoseph
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Panger, Gary R.
 ; APPLICANT: Smith, Carole L.
 ; APPLICANT: Durham, Margarita
 ; APPLICANT: Scolk, John A.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.563
 ; CURRENT APPLICATION NUMBER: US/10/066,543
 ; CURRENT FILING DATE: 2002-01-31
 ; NUMBER OF SEQ ID NOS: 3417
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3022
 ; LENGTH: 477
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; ORGANISM: Homo sapiens
 US-10-066-543-3022

Query Match 14.6%; Score 185.4; DB 14; Length 477;
 Best Local Similarity 73.2%; Pred.No. 7.6e-40;
 Matches 301; Conservative 0; Mismatches 86; Indels 24; Gaps 4;

Qy	265	TACCGAAGCATGTCCCGTGGTGAGCTCTAGAGGGGGCGCATCAATCACATGACA	324
Db	74	TACAGAGGCGTGTGTCGCGTGGCGAGGTAAAGA-GGGTCAGTATGGTCAAGTGACA	132
Qy	325	GTGTGGTACTCTGGCAAGACAGTGATGTTTCAGAA-----TATCTAA	367
Db	133	GTGTGGTAACTCTGGCAAGACAGTGATGTTTAAGAAGGTTTCATAGTTTAAGAATTAATCTAA	192
Qy	368	AATAGTTTAAAACTGTAAAGCGCAGCGTGTATCTACACCCAGTTACTAGAAAACG	427
Db	193	AATATTTTAAAACTATAAAGCTGCAACATGATTTTACCTTAGTTACTAGAAAAC	252
Qy	428	AAGGGAAGCACTAGTCAGCTGAGTAAAGGAGGTGAAAACAGGAAGCACTTCTACTATC	487
Db	253	AAGGAAGCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAAT	307
Qy	488	TACAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTCAGACATATT	547
Db	308	CGACAAAAAACCTTCTAATGCATTATCAGAAAAAGATTTTATAACAAG-GAGGCATATT	366
Qy	548	GTCGTGTTAAGAAGGGGTCCTAAAGAAAAGCACTTGCTAAGTTAGCAACTGTGAGGATGG	607
Db	367	GCTCAGTCAGAGGGGTTCTATAGAAGAAGCACTTACTAAGTTAGCGACTAACAGAACAA	426
Qy	608	CCAGTTTAAATATGAGCTCAAGCCCATCTGGGAGGGGACAGCAGGGGGA	658
Db	427	CCAGTTTAAAGATGAATTAATGCCAATTTGGGAGGCATGGCAGGTGTA	477

RESULT 15
 US-09-960-352-7612/c
 ; Sequence 7612, Application US/09960352
 ; Patent No. US20020137139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Wesley C.
 ; APPLICANT: Tao, Nengbing
 ; APPLICANT: Byatt, John C.
 ; APPLICANT: Machialagan, Nagappan
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 ; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
 ; FILE REFERENCE: 16511.006/37-21(10298)C
 ; CURRENT APPLICATION NUMBER: US/09/960,352
 ; CURRENT FILING DATE: 2001-09-24
 ; NUMBER OF SEQ ID NOS: 15112
 ; SEQ ID NO 7612
 ; LENGTH: 432
 ; TYPE: DNA
 ; ORGANISM: Bos taurus

[illegible]

```

; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (741)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-205-258-232

Query Match      3.7%; Score 46.8; DB 4; Length 760;
Best Local Similarity 57.1%; Pred. No. 0.0032;
Matches 84; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 TTTTNTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTATGATCAAGG 60
Db TTTTNTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTATGATCAAGG 689

Qy 61 ATTATTAAAGTCATACATGCAAAACATCTGTAAGTCAATGCAAAAGATCAATGTAA 120
Db GTGTTTATTATTATACCTACAAAGAAACAAAGATGATGTTATCAAAAGGACAAATTAC 629

Qy 121 AAACACTCCCAAAATCTGCAACTGTCA 147
Db 628 AAACATAAGATAGTAACATAGCTTTCA 602

RESULT 7
US-09-734-675-3/c
; Sequence 3, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C1000862
; CURRENT APPLICATION NUMBER: US/09/734, 675
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 38844
; TYPE: DNA
; ORGANISM: Human
; US-09-734-675-3

Query Match      3.6%; Score 45.6; DB 4; Length 38844;
Best Local Similarity 55.8%; Pred. No. 0.036;
Matches 87; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 2 TTTTNTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTATGATCAAGG 61
Db 24608 TCTCTTCTATTTCGAAATTCACATAAATCTTTTTTTTTTTTTTTTGGTTAAAGG 24549

Qy 62 TTTTATTAGTCATACATGCAAAACATCTGTAAGTCAATGCAAAAGATCAATGTAA 121
Db 24548 TTTTAAATGAAAAACCTTGAAACCTCTTGCATAATGATATTATAATGCTTTTAAAC 24489

Qy 122 AACACTCCCAAAATCTGCAACTGTCAATGTCAAAAAA 157
Db 24488 ACAGTCTCTAAATAGTAAATTTGCTTTTITAAAGAA 24453

RESULT 8
US-09-370-253-1/c
; Sequence 1, Application US/09370253
; Patent No. 6165792
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: BB-1200
```

```

; CURRENT APPLICATION NUMBER: US/09/370,253
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: 60/097,222
; EARLIER FILING DATE: August 20, 1998
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 1882
; TYPE: DNA
; ORGANISM: Hordeum vulgare
; US-09-370-253-1

Query Match      3.6%; Score 45; DB 3; Length 1882;
Best Local Similarity 54.5%; Pred. No. 0.015;
Matches 90; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1 TTTTNTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTATGATCAAGG 60
Db 1860 TTTTNTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTATGATCAAGG 1801

Qy 61 ATTATTAAAGTCATACATGCAAAACATCTGTAAGTCAATGCAAAAGATCAATGTAA 120
Db 1800 TTTTNTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTATGATCAAGG 1741

Qy 121 AAACACTCCCAAAATCTGCAACTGTCAATGTCAAAAAAGTTGTTC 165
Db 1740 GAACAATCGAATGATGATCTTGGCACTTAATTGATATTTCTC 1696

RESULT 9
US-08-924-747-25/c
; Sequence 25, Application US/08924747
; Patent No. 6063570
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,747
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: SOYBEAN
```

Query Match 3.4%; Score 43.2; DB 4; Length 615;

This Page Blank (uspto)

QY	623	ACTCAAGCCCATCTGGGAGGACACACGAGGGGAGGGGGCTCAAG-----	671
Db	1674	ACTCAAGCCCATCTGGGAGGACGCGAGTGAGGGTGGGGCGCGAAGGAGTGC	1615
QY	672	-AGAGACACTGATAAGATCGGCCATTTGTCTACTCTGTTTGACAGAAATTAACCGTTAA	730
Db	1614	AAGAGACACTGATAAGATCGGCCATTTGTCTACTCTGTTTGACAGAAATTAACCGTTAA	1555
QY	731	AAAGCTTTACCCGTGACACTTTTATTCAGTTGAATTAATCTCATGTAACAATGAGTGTAA	790
Db	1554	AAAGCTTTACCCGTGACACTTTTATTCAGTTGAATTAATCTCATGTAACAATGAGTGTAA	1495
QY	791	TTAATCTCTACTTCAATATTAGTCAAAATCTGCTGCTCTCTTTGATGACGTCGTGTTTC	850
Db	1494	GTAATCTCTACTTCAATATTAGTCAAAATCTGCTGCTCTCTTTGATGACGTCGTGTTTC	1436
QY	851	ACACACTCCACCCAGCACCCACGACTAGGACAGAAATCTCTGTTAGAGGCAACACAG	910
Db	1435	-CACACTCCACCCAGCACCCACGACTAGGACAGAAATCTCTGTTAGAGGCAACACAG	1377
QY	911	GAGCGAGTCTGTTCAAGCCCTGAGAACCCGCTGAGTGTGATTTTATGAGAACTCAC	970
Db	1376	GAACCGAGTCTGTTCAAGCCCTGAGAACCCGCTGAGTGTGATTTTATGAGAACTCAC	1317
QY	971	TATGAATCAAGAGCAGACGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACG	1027
Db	1316	TATGAATCAAGAGCAGCTGAGTGTGATTTTATGAGAACTCAC	1257
QY	1028	TAATGAGCTGGCTGATAGTTACAGTTCGCTTACATGCGAGCGTGTCTATTAAGGAGGC	1087
Db	1256	TAATGAACATGGCTGATAGTTACAGTTCGCTTACATGCGAGCGTGTCTATTAAGGAGGC	1197
QY	1088	TGTCGTGTCTACACGGTCTGGAGCTGAGGAGGTCTGCAACCCCTGAGCCGAGAGCT	1147
Db	1196	TGTCGTGTCTACACGGTCTGGAGCTGAGGAGGTCTGCAACCCCTGAGCCGAGAGCT	1137
QY	1148	GCAGTCTCTTAAAGGCAAAA---GTCTCTCAACAGCTTAGTGTCTTACGTTCTTCAGCAC	1204
Db	1136	GGGTCTCTTAAAGGCAAAA---GTCTCTCAACAGCTTAGTGTCTTACGTTCTTCAGCAC	1077
QY	1205	AACGCAACTTAGTTCACAGGATTTTGGCAATTTCTTAACTCTGAGCAAGATAGGGGATT	1264
Db	1076	AACGCAACTTAGTTCACAGGATTTTGGCAATTTCTTAACTCTGAGCAAGATAGGGGATT	1017
QY	1265	TT 1266	
Db	1016	TT 1015	
RESULT 2			
AK081613/c			
LOCUS			
DEFINITION	AK081613	1796 bp mRNA linear HTC 05-DEC-2002	
		Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched	
		library, clone: C130051H24 product: RAS-related C3 botulinum	
		substrate 1, full insert sequence.	
ACCESSION	AK081613		
VERSION	AK081613.1	GI:26349264	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to		

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamahata, I.,
Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Bash, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, C., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, J., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6921), 685-690 (2001)
21095660
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1796)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan [E-mail: genome-res@gs.c.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216]
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

COMMENT

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES
sou

```

FEATURES             Location/Qualifiers
     1..1796
         /organism="Mus musculus"
         /mol_type="mRNA"
         /strain="C57BL/6J"
         /db_xref="FANTOM_DB:C130051H24"
         /db_xref="taxon:10090"
         /clone="Cl30051H24"
         /tissue_type="head"
         /clone_lib="RIKEN full-length enriched mouse cDNA library"
         /dev_stage="16 days embryo"
     <1..304
         /note="unnamed protein product; RAS-related C3 botulinum
         substrate 1 [MGDB|MGI:97845, GB|AK011072, evidence: BLASTN,
         100%, match=987]
         putative"
         /codon_start=2
         /protein_id="BAC38272.1"
         /db_xref="GI:26349285"
         /translation="VRAKWYPEVRHHCNPTPIILVGYKLDLRDDKDTIEKLKKKKLTP
         ITPQGLAMAKETGAVKYLECALSATORGLKTVFDEAIRVLCPPVKKRKRKKLL"
         1781..1786
         /note="putative"
     1796
         /note="putative"
     470 a          413 c          371 g          542 t
BASE COUNT
ORIGIN

```

Query Match	67.2%;	Score 850.2;	DB 11;	Length 1796;
Best Local Similarity	85.6%;	Prod. No. 2e-102;		
Matches 1080;	Conservative 0;	Mismatches 128;	Indels 53;	Gaps 10;
QY	52	GATTCAAGGATTTTAAGTCATACATGCAAAACATACGTCTAACTGCATTAGCAAAAGA	111	
Db	1796	GATTCAGGATTTTAAATTCATACATGCAAAACATACGTCTAACTGCATTAGCAAAAGA	1737	
QY	112	TCAATGTAAAAACACTCCACAATTTCTGCACTGTCAATTGTAAGAAAAGTTTGTCTTAGTGG	171	
Db	1736	TCAATGTAAAAACACTCCACAATTTCTGCACTGTCAATTGTAAGAAAAGTTTGTCTTAGTGG	167	
QY	172	TCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTAGTTGTTACAGAAACGGCTTAGCA	231	
Db	1676	TTGAAAGGCCCAACACTGCAATTTCTGCAGTGAGTTAGTTGTTACAGGACAGCGTTAGCA	1617	
QY	232	CTAGCGTGTGACAGAACTCTACAGACCCAAAGG-----TACCGG	270	
Db	1616	CTAGCAGTTGACAGAACTCTACAGACCCCAAGGAACATCTCTAGGACAGGCGCATGACAGG	1557	
QY	271	AAGCATGTGTCGGTGGGTGAGTGTAGAGGGGGCGCATCAATCACATCACAGTGTGTG	330	
Db	1556	AAGCGTCTGTCCACGTGGCGAGGTCTAGA--GGCAGCATTAGTCACATGACAGTGTGTG	1500	
QY	331	GTACTCTGGCAAGACAGTGTGTTTCAGAAATCTTAAATAGTTTAAAAACTGTAAAGCC	390	
Db	1499	GTACTGTTTGAAGTTTCATAGTTTAAAGATATCTTAAATAGTTTAAAAACCGTAAAGCT	1440	
QY	391	GCAGCACGTGATTTCTACACCCAGTTACTAGAAACGAAGGAGCAGCTAGT-CAGCTGA	449	
Db	1439	GCAACACATGATTTGCAACACCTAGTTGTGTAGAAACTAAGGAAGCACTAATCTAGCTCTGA	1380	
QY	450	GTAAAGGAAGGTGAAAAACAGGAACGCACTTCTACTATCTACCAAAAAAATCTCCGAA---	506	
Db	1379	GTAAAGTAGGCGAAAAACAGGACGCACTTCTACTGAGCTACCAAAAAAACAACAAAAAT	1320	
QY	507	---TGCATTATCAGAAAGATCTTATAGTACAGGTCAGACATATTCGTCGTTAAGAGGGG	563	
Db	1319	CTCTCGGTTATCAGAAAGATCTTATAGTACAGGTCAGACATATTCGTCGTTAAGAA--GGG	1261	
QY	564	GTCTCTAAAGAAAGACATTCGTTAAGTTAGCAACTGTGAGGATGGCCAGTTTAAATATGGA	623	

1260	Db	GTCTAAAGAAAGCACTTCCTAAGTTAG-AACTGTGAGGGTGGCCAGCTTTAAATATGGA	1205
624	Qy	CTCAAGCCCCCATCTCTGGGAGGACAGCAGGGGGAAGGGGGCTCAAG-----	671
1201	Db	CTCAAGCCCCCATCTCTGGGAGGACAGCAGTGTGAGGGTGGGGGCGCGAAGGATGCTCA	1142
672	Qy	AGAGACACTGATTAAGATCGGCCAATTGTCTACTCTACTCTTTGACAGAAATTAACCGTTAAA	731
1141	Db	AGAGACACTGATTAAGATCGGCCAATTGTCTACTCTCTTTGACAGAAATTAACCGTTAAA	1082
732	Qy	AAGCTTTACCGGTGACACTTTTATTCAAGTTGAATTAATCCATGTACAAATGTAGTGTAAAT	791
1081	Db	AAGCTTTACCGGTGACACTTTTATTCAAGTTGAATTAATCCATGTACAAATGTAGTGTAAAG	1022
792	Qy	TAATCTCTACTTCATATTAGTCAAAAATACCTGTCTCTCTCTTTGATCAAGCTCGTGTTC	851
1021	Db	TAATCTCTACTTCATATTAGTCAAAAATACCTGTCTCTCTCTTTGATCATGTTCGTGTT--	964
852	Qy	CACACTCCAGCCAGCACACCCACGACTTAGGAACAGAATACTTCTGTAGAGCCACACAGG	911
963	Db	CACACTCCAGCCAGCACACCCACAACTAGGAACAGAATACTTCTATTAGAGCCACACAGG	904
912	Qy	AGCCAGAGTTCCTGTTCAAGCCCTGCAGAGCCGCTCAGCTGGTATTTTAGAGAATCACT	971
903	Db	AACAGAGTTCCTGTTTCAAGTCTGCAAAAGCTAGTCGGCTGGTCTTTTAGAAAATCACT	844
972	Qy	ATGAAATCAAGAGCAGAGCTGTTTACACCCATC---GTGACGTACAGTACAAAGATTACGT	1028
843	Db	ATGAAATCAAGAGCTGAGCTGTACACTCATCACTGTGAGGTACAGTACAAAGATTACGT	784
1029	Qy	AATGAGCATGGCTGTATAAGTTACAGTGGGTTCATCGGACGCGTGTCATTAAAGAGGCT	1088
783	Db	AATGAAATCATGGCTGTATAAGTTACAGTGGCTTCATCATGGCAACGTGTCAITTAAGAGGCT	724
1089	Qy	GTGCTGTGTACAGGCTCTGGGAGCTACGGAGGGTCTGCACCCCTGAGCCACGAGCTG	1148
723	Db	GTGCTGTGTACAGCGGTCTGGGAATCTTGGAAAGGCTCTGCAACCCCTGGCTCCAAAGCTG	664
1149	Qy	CAGTCTCTCTTAAGCACAAA---GTCCTCAACAGCTTAGTCTCTACGTGTTCTCAGCAC	1205
663	Db	CGGTCTCTCTTAGCAACAGAAGTCTCTCAAGAGGTTTAGTGCTTAGGTGTTCTCAGCAC	604
1206	Qy	ACGCAACTTAGTTCACAGGTATTTTGGCAATTCCTTAATCTGAGCAAGAAATGAGGATTT	1265
603	Db	ACGCAACTTAGTTCAGAGGTATTTTGGCAATTTTAACTCTGAGCAAGAAATGAGGATTT	544
1266	Qy	T 1266	
543	Db	T 543	

RESULT 3

AK047969/c

LOCUS

DEFINITION

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ACCESSION

VERSION

KEYWORDS .

SOURCE

ORGANISM

REFERENCE

REFERENCE

AUTHORS
TITLE

FILE
JOURNALJOURNAL
MEDLINEMEDLINE
PUBMED

REFERENCE

REFERENCES

0
0
0
0
0
0
0

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

204939374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

20530913
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Schiavi, P.,
Quackenbush, J., Schriml, L. M., Staib, L. J., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6921), 695-699 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

21085660
11217851
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL
REFERENCE
AUTHORS

6 (bases 1 to 2270)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE
JOURNAL

Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES
source

1. 2270
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C130025F01"
/db_xref="taxon:10090"
/clone="C130025F01"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
200..778
/notes="unnamed protein product; PAS-related C3 botulinum
substrate 1 (MGB|MG1:97845, GB|AK011072, evidence: BLASTN,
100%, match=987)
putative"
/codon_start=1
/protein_id="BAC33203.1"
/db_xref="GI:26339064"
/translation="MQAIKCVVVGDAVGKTCLLISYTTNAPFGEVPTVFDNYSANV
YDGRFVNGLMDTAGQEDYDLRLPLSYFQTDVFLICFLSVSPASEFNRKRYPEVR
HHCPTPIILVGTGLDRDDKDTIEKLEKLTPTIYQGLAMAKBIGAVKYLECSAL
TORGKLVDFEAIKRAVLCPPVKKRKKCLLL"
2254..2259
/note="putative"
2270
/note="putative"

CDS

554 a 569 c 644 t
BASE COUNT
ORIGIN
Query Match 67.1%; Score 849.6; DB 11; Length 2270;
Best Local Similarity 85.6%; Pred. No. 2.1e-102;
Matches 1079; Conservative 0; Mismatches 129; Indels 52; Gaps 10;
QY 52 GATTCAAGGATTATTATTAAGTCATACATGCAAAACATACATGCTGCTAAGTCAATGCAAAAGATTTTCTAGTGG 111
DB 2269 GATTCAAGGATTATTATTAATTCATACATGCAAAACATACATGCTGCTAAGTCAATGCAAAAGATTTTCTAGTGG 2210
QY 112 TCAATGTAACACATCCCAATTTGCAACTGTCATTAATGCAAAAGATTTTCTAGTGG 171
DB 2209 TCAATGTAACACATCCCAATTTGCAACTGTCATTAATGCAAAAGATTTTCTAGTGG 2150
QY 172 TCGAAAGGCCCAACACTGCTGTTCTGCCAGTGAAGTGTGTTACAGACGGCGTTAGCA 231
DB 2149 TTGAAGGCCCAACACTGCTGTTCTGCCAGTGAAGTGTGTTACAGACGGCGTTAGCA 2090
QY 232 CTAGCGCTTTACAGAACCTTCACAGACCCCAAGG-----TACCAG 270
DB 2089 CTAGCGCTTTACAGAACCTTCACAGACCCCAAGGACATCTCTAGGACAGCCATGACAGG 2030
QY 271 AGCATGTGTCGGCTGGTGGTGGTGTAGAGGGGGCGGCATCAATCAATGACAGTGTG 330
DB 2029 AAGCGCTGTCACGCTGGCGAGGCTCTAG--GGCAGCATTTAGTGCATGACAGTGTG 1973
QY 331 GTACTCTGCAAGACAGTGAATGTTTTCAGATATCTAAATAGTTTAAACTGTAAAGCC 390
DB 1972 GTACTCTGCAAGGTTTCATAGTTTAAAGATATCTAAATAGTTTAAACTGTAAAGCC 1913
QY 391 GCAGCAGCTGATTCTTACACCCAGGTTTACAGAAACGAGGAGGACCTAGT-CAGCTCA 449
DB 1912 GCACACATGATTTCACACCTAGTTGCTAGAAACTAGGAGGACCTAGTCTCTGTA 1853
QY 450 GTAAGGAGGTGAAACAGGACCCATCTTCTATCTATCTACCAAAAAAATCTCCGAA--- 506
DB 1852 GTAAAGTAAAGCGCAAAACAGGACGCACTTCTACTGAGCTACCAAAAAAATCTCCGAA--- 1793
QY 507 --TGCAATTATCAGAAAGATCTTATAGTACAGTGTGACATATTTGCTCTAGAGGGGG 564
DB 1792 TCTGCGTTATCAGAAAGATCTTATAGTACAGGTCAGATATTTGCTCTAGAGG--GGGG 1734

QY 565 TCCTAAGAAAGACCTTCTAAGCTTACCACTCTGAGATGGCGAGTTTAAATATGAC 624
 Db 1733 TTCTAAGAAAGACCTTCTAAGCTTACCACTCTGAGATGGCGAGTTTAAATATGAC 1675

QY 625 TCAACGCCCATCTGGGAGGAGCAGCAGGAGGGGAGGGGGCTCAAG-----A 672
 Db 1674 TCAACGCCCATCTGGGAGGAGCAGCAGGAGGGGAGGGGGCTCAAGGGATGCTCAA 1615

QY 673 GAGACACTGATAAGATCGCCATTTGTTCATCTACTGTTTGTGACAGAAATTAACCGTTAAAA 732
 Db 1614 GAGACACTGATAAGATCGCCATTTGTTCATCTACTGTTTGTGACAGAAATTAACCGTTAAAA 1555

QY 733 AGCTTTACCGCTGACACTTTTATTCAGTTGAAATTAATCTCATGTAACAATGTAAGTGAATTT 792
 Db 1554 AGCTTTACCGCTGACACTTTTATTCAGTTGAAATTAATCTCATGTAACAATGTAAGTGAATTT 1495

QY 793 AATCTCTACTTCAATATTAGTCAAAATACCTGTCTCTCTTTGTATGACGTCGTGTTTCAC 852
 Db 1494 AATCTCTACTTCAATATTAGTCAAAATACCTGTCTCTCTTTGTATGACGTCGTGTTTCAC 1437

QY 853 ACATCTCCACCCAGCAGCAGCCAGCTAGGACAGAAATCTTCGTTAGAGGCAACACAGGA 912
 Db 1436 ACATCTCCACCCAGCAGCAGCCAGCTAGGACAGAAATCTTCATTTAGAGGCAACACAGGA 1377

QY 913 GCCAGAGTCTGTTCAAGCGCTGCAAGACCGCTGAGCTGTGATTTTGTAGAGAACTCACTA 972
 Db 1376 ACCAGAGTCTGTTCAAGCGCTGCAAGACCGCTGAGCTGTGATTTTGTAGAGAACTCACTA 1317

QY 973 TGAATCAAGAGCAGAGCTGTTACACCCCATC- --GTGACGTACAGTACAAAGTTACGTA 1029
 Db 1316 TGAATCAAGAGCAGAGCTGTTACACCCCATCCTCATCTGTGAGCTACAGTACAAAGTTACGTA 1257

QY 1030 ATGACATGGGCTGATAGTTACAGTGGTTACATGGCAGCGTGTCTATTAGAGGAGCTG 1089
 Db 1256 ATGACATGGGCTGATAGTTACAGTGGTTACATGGCAGCGTGTCTATTAGAGGAGCTG 1197

QY 1090 TGCTGTGTCACAGCTGTGGAGCTACGGAGGCTGTGACCCCTTGAGCCCAAGAGCTGC 1149
 Db 1196 TGCTGTGTCACAGCTGTGGAGCTGTGAGGAGCTGTGAGGAGCTGTGAGGAGCTGC 1137

QY 1150 AGTCTTCTTAAAGACAAA- --GTCTCTCAACAGCTAGTGTCTTACGTTCTTCAGCACAA 1206
 Db 1136 GTCTTCTTAAAGACAAA- --GTCTCTCAACAGCTAGTGTCTTACGTTCTTCAGCACAA 1077

QY 1207 CGCACTTAGTTCACAGTATTTTGGCAATTTTAACTGAGCAAGAAATAGGGATTTT 1266
 Db 1076 CGCACTTAGTTCACAGTATTTTGGCAATTTTAACTGAGCAAGAAATAGGGATTTT 1017

RESULT 4
 AK076023/c
 LOCUS 2308 bp mRNA linear HTC 07-DEC-2002
 DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 enriched library, clone:2610100016 product:RAS-related C3 botulinum
 subunit 1, full insert sequence.

ACCESSION AK076023
 VERSION AK076023.1 GI:26344957
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

20499374
 11042159
 3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed capillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

20530913
 11076861
 4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schriml, L.M., Stauble, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Hill, D.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
 Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wyshak-
 Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S.
 and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)

21085660
 11217851
 5

The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6 (bases 1 to 2308)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
 Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
 Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
 Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D.,
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
 Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
 Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A.,
 Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

source	Location/Qualifiers
1. 2308	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM DB:2610100016"
	/db_xref="MGI:1898616"
	/db_xref="taxon:10090"
	/clone="2610100016"
	/tissue type="whole body"
	/clone lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="10 days embryo"
239..817	
	/notes="unnamed protein product; RAS-related C3 botulinum substrate 1 (MG MG1:97845, GB AK011072, evidence: BLASTN, 100%, match=987)"
	putative"
	/codon_start=1
	/protein_id="BAC36128.1"
	/db_xref="GI:26344958"
	/db_xref="MGI:97845"
	/translation="MQAIKVVGVGAVGKTCLLISYTTNAPGEVPTVFQNSAVN MVDGKPNLGLMDTAGQEDYRLPLSPQTDVFLICFSLVSAASENVRKWPYEV RHCNPPIILVGTPLKLRDDKDTIEKLKELKLPYIPQGLAWAKEIGAVKYLECSAL TORGLKTVFDEAIRAVLCPPPVKKRKKCLLL"
2293..2298	
	/notes="putative"
2308	
	/notes="putative"
BASE COUNT	559 a 580 c 518 g 650 t 1 others
ORIGIN	
Query Match	64.4%; Score 815.4; DB 11; Length 2308;
Best Local Similarity	84.9%; Pred. No. 6.1e-98;
Matches 1071; Conservative	0; Mismatches 136; Indels 54; Gaps 12;
QY	52 GATTCAGGATTTTAAAGTCATACATGCAAAACATACCTGCTAACTGCATTAGCAAAAGA 111
DB	2308 GATTCAGGATTTTAAATTTTACATATGCAAAACATACCTGCTAACTGCATTAGCAAAAGA 2249
QY	112 TCAATGTAAACACATCCCAATTCGCACTGCTCAATTCGAAATAGTTTGTCTAGTGG 171
DB	2248 TCGATGTAAACACATCCCAATTCGCACTGCTCAATTCGAAATAGTTTGTCTAGTGG 2189
QY	172 TCGAAGGCCCAACATCTGTGTTTTCGCCAGTGAAGTTAGTTGTACAGAACGGCGTTAGCA 231
DB	2188 TTGAGAGGCCCAACATCTGTGTTTTCGCCAGTGAAGTTAGTTGTACAGAACGGCGTTA-CA 2130
QY	232 CTAGCGCTTGACAGACCTCAGACCCCAAGG-----TACCGG 270
DB	2129 CTAGCAGTTGACAGAACCTTCAGACCCCAAGGACATCTCTAGGCAGACCCATGACAGG 2070
QY	271 AAGCATGTCTCCGCTGGGTGAGGTCTAGAGGGGGGCGCATCAATCACATGACAGTGTG 330
DB	2069 GAGCGTCTGTCACGTGGCGGAGGTCTAGA---GGCAGCATTAGTCACATGACAGTGTG 2013
QY	331 GTACTCTGGCAAGACAGTGAATGTTTTCAGAAATCTCTAAATAGTTTAAAACTGTAAAGCC 390
DB	2012 GTACTCTGTTAGAGGTTTCATAGTTTAAAGAAATCTCTAAATAGTTTAAAACTGTAAAGT 1953
QY	391 CGACGACGTGATTTCTACACCCAGTACTAGAAACGAAGGAGGAGCACTAGT-CAGCTGA 449
DB	1952 GCAACACATGATTTGACACACCTAGTTGCTAGGAACTAAGGAAGCACTAAGTCTCTGA 1893
QY	450 GTAAAGGAGGTGAAACAGGACCGCACTTCTACTATCTACCAAAAATCTCCGAA---506
DB	1892 GTAAAGTAAAGCGAACAACGAGCGCACTTCTACTGAGCTTACCAAAAAAACAACAAATC 1833
QY	507 --TGCAATTATCAGAAGATCTTATAGTACAGGTCAGACATATGCTCGTTAAGAGGGGG 564

Db	1832 TCTCGGTTATCAGAAAGATCTTATAGTACAGGTGACACATATTCCTCTGTTAAGAA-GGGG 1774
QY	565 TCCTAAAGAAAAGCAGCTTGTCTAGTAACTGTGAGGATGGCCAGTTTAAATATGAC 624
Db	1773 TTCTAAAGAAAAGCAGCTTGTCTAGTAACTGTGAGGATGGCCAGTTTAAATATGAC 1715
QY	625 TCAACGCCCATCTCTGGGAGGAGCAGCAGGGGGAAGGGGGCTCAAG-----A 672
Db	1714 TCAACGCCCATCTCTGGGAGGAGCAGCAGGTGAGGGTGGGGGGCGGAGAGGATGCTCAA 1655
QY	673 GAGACACTGATAAGATCGCCATTTGTCATCTACTGTTTACAGAAATTAACCGTTAAAA 732
Db	1654 GAGACACTGATAAGATCGCCATTTGTCATCTACTGTTTACAGAAATTAACCGTTAAAA 1595
QY	733 AGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCAATGTACAAATGTAGTGAAT 792
Db	1594 AGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCAATGTACAAATGTAGTGAAT 1535
QY	793 AATCTCTACTTCAATATTAGTCAAAATCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 852
Db	1534 AATCTCTACTTCAATATTAGTCAAAATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1477
QY	853 ACATCTCACCCAGCAGCACCCAGCAGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 912
Db	1476 ACATCTCACCCAGCAGCACCCAGCAGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1417
QY	913 GCAGAGTTCTGTCTCAAAAGCCTGAGAGGCGGTGAGTGGTATTTTAGAGAACTCACTA 972
Db	1416 ACCAGAGTTCTGTCTCAAAAGCCTGAGAGGCGGTGAGTGGTATTTTAGAGAACTCACTA 1357
QY	973 TGAATCAAGAGCAGAGCTGTACACCATC---GTGAGCTACAGTACAAAGTTAGTA 1029
Db	1356 TGAATCAAGAGCAGCTGAGCTGTACACCATC---GTGAGCTACAGTACAAAGTTAGTA 1297
QY	1030 ATGAGCATGGCTGATAAGTTACAGGTGCTTACATGGCGAGCGTGTCAATTAAGAGGCTG 1089
Db	1296 ATGAGCATGGCTGATAAGTTACAGGTGCTTACATGGCGAGCGTGTCAATTAAGAGGCTG 1237
QY	1090 TGTGTGTCTACAGCGTCTGGGAGCTACGGGAGGCTGTGACCCCT--GAGCCCAAGAGCTG 1148
Db	1236 TGTGTGTCTACAGCGTCTGGGAGCTACGGGAGGCTGTGACCCCTGTGACAGAGCTG 1177
QY	1149 CAGTCTCTTAGGACAAA---GTCTCTCAACAGCTTAGTGTCTAGTGTCTCAGACAC 1205
Db	1176 CGGTCTCTTAGGACAAA---GTCTCTCAACAGAGTCTCTCAAGAGAGTTAGTGTCTCTCAGAC 1117
QY	1206 AGCAACTTAGTTCAGAGGTATTTTGCAATTTCTAATCTGAGCAAGATAGGGGATTT 1265
Db	1116 AGCCAACTAGTTCAGAGGTATTTTGCAATTTCTAATCTGAGCAAGATAGGGGATTT 1057
QY	1266 T 1266
Db	1056 T 1056
RESULT 5	
BQ191985	
LOCUS	BQ191985 760 bp mRNA linear EST 30-APR-2002
DEFINITION	UI-R-DR1-cky-h-11-0-UI-s1 UI-R-DR1 Rattus norvegicus cDNA clone
ACCESSION	UI-R-DR1-cky-h-11-0-UI 3', mRNA sequence.
VERSION	BQ191985
KEYWORDS	BQ191985.1 GI:20367536
SOURCE	EST.
ORGANISM	Rattus norvegicus (Norway rat)
REFERENCE	1 (bases 1 to 760)
AUTHORS	Ronald, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)

QY	573	AAAAGCACTTCTAGTTAGCAACTGTGAGGATGCCAGTTTAAATATGGACTCAACGCC	601
Db	542	AAAAGCACTTCTAGTTAGTTAGCAACTGTGAGGATGCCAGTTTAAATATGGACTCAACGCC	601
QY	633	CCATCTGGGAGGACACGACGAGGGGAGGGGGCTCAAG	671
Db	602	CCATCTGGGAGGACACGACGAGGGGAGGGGGCTCAAG	640
RESULT 9			
CA339407/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
QY	113	CAATGTAAAACACTCCACAAATTCTGCACTGCAATTGAAAAAGATTGTTCTAGTGGT	172
Db	644	CAATGTAAAACACTCCACAAATTCTGCACTGCAATTGAAAAAGATTGTTCTAGTGGT	585
QY	173	CGAAAGGCCCAACACTGTGTTCTTGCAGTGAGTTAGTTGTACAGAACGGCGTTAGCAC	232
Db	584	CGAAAGGCCCAACACTGTGTTCTTGCAGTGAGTTAGTTGTACAGAACGGCGTTAGCAC	525
QY	233	TAGCGCTTGACGAACCTTCACAGACCCAAAGGTACCGAAGCAGTGTGTCGCGTGGGTGA	292
Db	524	TAGCGCTTGACGAACCTTCACAGACCCAAAGGTACCGAAGCAGTGTGTCGCGTGGGTGA	465
QY	293	GGTCTAGAGGGCGGCATCAATCAATGACAGTGTGGTACTCTGCGAAGACAGTGTATG	352


```

/db xref="taxon:10090"
/clone="IMAGE:6511045"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 134"
/notes="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dT. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT      249 a   224 c   201 g   301 t      3 others
ORIGIN
Query Match      47.3%; Score 599.2; DB 13; Length 978;
Best Local Similarity 85.9%; Pred. No. 1.9e-69;
Matches 767; Conservative 0; Mismatches 94; Indels 32; Gaps 8;

QY 402 TTCTTACCCAGTTACTAGAAAACGAGGAGCACTAGTCTAG--CTGAGTAAAGGAAG 459
Db   |||
QY 953 TTGCACCACTAGTTGTTAGAAANCTAAGGAAGCCACTAAGTCTGAGTAAAGTAGG 894
Db   |||
QY 460 GTGAACAGGAGGACCTTCTACT--ATCTACAAAATAATCTCCGA-----TGCAT 511
Db   |||
QY 893 GCGAAACAGGAGGACCTTCTACTTGGCTACCAAAAAAACAATAATCTTGCCT 834
Db   |||
QY 512 TATCAGAAAGATCTTATAGTACAGGTGAGATATTTGCTGTTAAGAGGGGTCTTAA 571
Db   |||
QY 833 TATCAGAAAGATCTTATAGTACAGGTGAGATATTTGCTGTTAAGAGGGGTCTTAA 776
Db   |||
QY 572 GAAAGACCTTGTAAAGTATAGCACTGTGAGGATGCCAGTTTAAATATGACTCAACGC 631
Db   |||
QY 775 GAAAGACCTTGTAAAGTATAGCACTGTGAGGATGCCAGTTTAAATATGACTCAACGC 716
Db   |||
QY 632 CCATCTGGGGAGGACACAGGAGGAGGGGGCTCAAG-----AGAGACAC 679
Db   |||
QY 715 CCATCTGGGGAGGAGGACGAGGAGGAGGGGGCGGAGGAGTGTCTCAAGAGACAC 656
Db   |||
QY 680 TGATAGATCGGCCATTTGTCATCTACTGTTTACAGAAATTAACGTTAAAGAGCTTTA 739
Db   |||
QY 655 TGATAGATCGGCCATTTGTCATCTACTGTTTACAGAAATTAACGTTAAAGAGCTTTA 596
Db   |||
QY 740 CCGGTGACACTTTATTCAGTTGAATTAATCTCCATGTACAAATGTAGTGTAAATTAATCTCT 799
Db   |||
QY 595 CCGGTGACACTTTATTCAGTTGAATTAATCTCCATGTACAAATGTAGTGTAAATTAATCTCT 536
Db   |||
QY 800 ACTTCATATAGTCAAAATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 859
Db   |||
QY 535 ACTTCATATAGTCAAAATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 478
Db   |||
QY 860 ACCGAGCACCCACGACTAGGAACAGAAATACTTGTGTAGAGGACACAGAGGACGAG 919
Db   |||
QY 477 ACCGAGCACCCACGACTAGGAACAGAAATACTTGTGTAGAGGACACAGAGGACGAG 418
Db   |||
QY 920 TTCTGTTCAAAGCTGCAAGCCGGTCAAGTGGTATTTTAGAGAACTCACTATGAAATC 979
Db   |||
QY 417 TTCTGTTCAAAGCTGCAAGCCGGTCAAGTGGTATTTTAGAGAACTCACTATGAAATC 358
Db   |||
QY 980 AAAGAGCAGAGCTGTATACCCATC---GTGAGTACAGTACAAAGTACGTAATGAGCA 1036
Db   |||
QY 357 AAAGAGCAGAGCTGTATACCCATC---GTGAGTACAGTACAAAGTACGTAATGAGCA 298
Db   |||
QY 1037 TGGGCTGATAGTTACAGTGGTGTACATGGCAGGCTGTCATTAAGGAGGCTGTCTGTG 1096
Db   |||
QY 297 TGGGCTGATAGTTACAGTGGTGTACATGGCAGGCTGTCATTAAGGAGGCTGTCTGTG 238
Db   |||
QY 1097 TCACAGGTCTGGGAGCTACGGAGGGTCTGCACCCCTGAGCCCAAGAGCTCAGTCTTC 1156
Db   |||
QY 237 TCACAGGTCTGGGAGCTACGGAGGGTCTGCACCCCTGAGCCCAAGAGCTCAGTCTTC 178
Db   |||
QY 1157 TTAAGGACAAA---GTCTCTCAACAGCTTAGTCTTCTAGCTTTCTCAGCAACGCAACT 1213
Db   |||
QY 177 TTAGCAACAGAGAGTCTCTCAAGAGAGTGTAGTGTCTAGTGTCTCAGCAACGCAACT 118
Db   |||
QY 1214 TAGTTTCACAAGGATTTTGGCAATTTCTTAATCTGAGCAGAAATAGGGGATTTT 1266

```

Db 117 TAGTTCAGAAGGTAATTTGGCAATTTTAAATCTGAGCAAGAAATGGGGATTTT 65

RESULT 15

CA318577/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA318577.1 GI:24536701

EST.

Mus musculus (house mouse)

Mus musculus

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 826)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaabs@mail.nih.gov

Tissue procurement: Dr. Jim Lin, University of Iowa

CDNA library preparation: Dr. M. Bento Soares, University of Iowa

CDNA sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. .826

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6813639"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH BMAP FW0"

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;

Site 2: Not I; The library was constructed according

Bonaldi, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with EcoR I adaptor, digested with NotI and then cloned.

directionally into pYX-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is ACCGAGACAG. This library was created for the University

Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,

program coordinator."

BASE COUNT 226 a 195 c 162 g 242 t 1 others

ORIGIN

Query Match 44.8%; Score 567; DB 14; Length 826;

Best Local Similarity 89.7%; Pred. No. 3.3e-65;

Matches 682; Conservative 0; Mismatches 56; Indels 22; Gaps 6;

QY 525 TTATGATACAGTTCAGACATATTCCTGTTAAGAGGGGGTCTTAAAGAAAGCACTTGC 584

Db 826 TTATGATACAGTTCAGACATATTCCTGTTAAGAGGGGGTCTTAAAGAAAGCACTTGC 768

QY 585 TAAGTTACAACTGTGAGGATGGCCAGTTTAAATATGCACTCAACGCCCATCTCGGAG 644

Db 767 TAAGTTAG-AACTGTGAGGGGTGGCCAGTNTAAATATGCACTCAACGCCCATCTCGGAG 709

us-09-717-321a-2.rst

Qy	645	GGACACAGGGGAAAGGGGGCTCAAG-----AGAGACACTGATTAAGATCGGC	692
Db	708	GGACGGCAGTGGAGGGTGGGGGGCGGAAGAGATGCTCAAGAGACACTGATTAAGATCGGC	649
Qy	693	CATTGTGCTACTCTGTTTACAGAAATTAACCGTTAAAGAGCTTTACCGGTGACACTTTT	752
Db	648	CATTGTGCCTACTGTTTACAGAAATTAACCGTTAAAGAGCTTTACCGGTGACACTTTT	589
Qy	753	TATTCAAGTTGAAATTAAGTCCATGTACAAATGTAGTGTAAATTAATCTCTACTTCATATTAGT	812
Db	588	TATTCAAGTTGAAATTAAGTCCATGTACAAATGTAGTGTAAATTAATCTCTACTTCATATTAGT	529
Qy	813	CAAAAATACTGTCTGTCTCCCTTGATGACGTCGCTGTTTACACACTCCGCCAGCACACCC	872
Db	528	CAAAAATACTGTCTGTCTCCCTTGATGATGTCGCTGTTT--CACACTCCGCCAGCACACCC	471
Qy	873	ACGACTAGGAAACAGAAATACTTCGTTAGAGCAACACAGGAGCCAGAGTTCCTGTTCAAAGC	932
Db	470	ACAACTAGGAACAGAAATACTTCATTAGAGCAACACAGGAACCCAGAGTTCCTGTTCAAAGT	411
Qy	933	CTGCAGAAAGCGGTCAGCTGGTATTTTAGAGAACTCACTATGAATCAAAAGACAGAGCT	992
Db	410	CTGCAAAAGCTAGTCGGCTCGCTCTTTTAGAAAACTCACTATGAATCAAAAGAGCTGAGCT	351
Qy	993	GTTTACACCCATC--GTGAGGTACAGTACAAAGTTAGTAAATCAGACATGGCTGATAAGT	1049
Db	350	GTCACTCTCATCACTGTGAGCTGACGATACAAAGTTAGTAAATGAACATGGCTGTATAAGT	291
Qy	1050	TACAGTGCCTTTACATGGCACGCTGTCAITTAAGGAGCTGTGCTGTGTCAACGGTCTGG	1109
Db	290	TACAGTGCCTTTACATGGCAACGTGTCAITTAAGGAGCTGTGCTGTGTCAACGGTCTGG	231
Qy	1110	GAGCTACGGAGGGTCTGCAACCCCTGAGCCAGAGCTGCAAGTCTTCTTAAGACAAA--	1167
Db	230	GAACTCTGGAAGGGTCTGCAACCCCTGAGCTCCAGAGCTCGGGTCTCTTAGCAACAGAAG	171
Qy	1168	-GTCTCTCAACAGCTTGTAGTGCTTACGTGTTCTCAGCAACACGCAATTAAGTTCACAAGGT	1226
Db	170	TCCTCAAGAGAGTTTAGTGCTTAGTGTTCTCAGCAACACGCAACTAGTTCAGAAGGT	111
Qy	1227	ATTTTGGCAATCTTAATCTGAGCAAGAAATAGGGGATTTT	1266
Db	110	ATTTTGGCAATTTTAAATCTGAGCAAGAAATGGGGATTTT	71

Search completed: November 23, 2003, 15:50:01
Job time : 2984.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:43:52 ; Search time 818.065 Seconds
(without alignments)
10338.975 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348

Sequence: 1 tgaacatcactgtgtgcc.....ctttggaacttgaaaaaaa 348

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_nus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gsl:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	324	93.1	429	9	AI059212 UI-R-CI-1
C 2	323.4	92.9	493	10	BF554834 UI-R-E0-C
C 3	323.4	92.9	687	12	BM389059 UI-R-D20-
C 4	321.8	92.5	459	13	BQ780699 UI-R-PF0-

C 5	321.8	92.5	640	13	BQ780117
C 6	321.4	92.4	477	9	AW520555
C 7	321.4	92.4	489	9	AA859879
C 8	321.4	92.4	565	12	BI273986
C 9	321.4	92.4	760	13	BQ191985
C 10	320.8	92.2	472	13	BQ190077
C 11	320.8	92.2	615	10	BQ380448
C 12	320.8	92.2	639	10	BE111691
C 13	320.8	92.2	705	13	BQ200122
C 14	318.2	91.4	465	9	AA899757
C 15	317	91.1	405	10	BE104696
C 16	317	91.1	412	13	BQ780657
C 17	317	91.1	517	9	AA859940
C 18	317	91.1	636	13	BQ780658
C 19	317	91.1	638	12	BI283790
C 20	313.8	90.2	469	12	BI276611
C 21	312.4	89.8	557	10	BG665005
C 22	310.6	89.3	446	12	BI286743
C 23	305.4	87.8	463	10	BF548903
C 24	304.8	87.6	316	9	AI412434
C 25	300	86.2	316	9	AI409795
C 26	297.2	85.4	407	9	AI009011
C 27	297.2	85.4	455	9	AI012196
C 28	295.6	84.9	553	9	AA799542
C 29	293.6	84.4	427	9	AI178082
C 30	292	83.9	431	9	AI410870
C 31	285.6	82.1	292	9	AI598992
C 32	278	79.9	294	9	AI236740
C 33	276.8	79.5	522	14	C06842
C 34	273.8	78.7	277	10	BF420446
C 35	251.2	72.2	277	10	BF420436
C 36	246.4	70.8	404	10	BF549054
C 37	243.4	69.9	644	14	CA339407
C 38	238.4	68.5	256	9	AI234716
C 39	227.6	65.4	394	12	BM022196
C 40	226.6	65.1	322	13	BQ757481
C 41	226.6	65.1	324	13	BQ756887
C 42	226.6	65.1	546	12	BQ033391
C 43	226.6	65.1	655	10	B3996534
C 44	226	64.9	408	12	BI157798
C 45	225.8	64.9	404	12	BM502810

ALIGNMENTS

RESULT 1
AI059212/c
LOCUS
DEFINITION
UI-R-CI-lb-e-11-0-UI.s1 429 bp mRNA linear EST 11-FEB-1999
UI-R-CI-lb-e-11-0-UI 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AI059212.1 GI:3332989
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 429)
AUTHORS
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL
MEDLINE
PUBMED
COMMENT
AI059212
UI-R-CI-lb-e-11-0-UI.s1 429 bp mRNA linear EST 11-FEB-1999
UI-R-CI-lb-e-11-0-UI 3', mRNA sequence.
AI059212
GI:3332989
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 429)
Ronald, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics

Seq primer: M13 Forward.

Location/Qualifiers

FEATURES

source

1. .429

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-C1-1b-e-11-0-UI"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-C1"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C1 library is a subtracted library derived from the UI-R-C0 library, which is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)." 130 a 87 c 91 g 121 t

BASE COUNT

ORIGIN

Query Match 93.1%; Score 324; DB 9; Length 429;

Best Local Similarity 95.7%; Pred. No. 4.5e-37;

Matches 333; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 TGAACATCACTGTCTTCCAGAGTACCAACTGTCATGTGATGATCGCGCCCTCT 60

Db 357 TGAACATCACTGTCTTCCAGAGTACCAACTGTCATGTGATGATCGCGCCCTAT 298

QY 61 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTGGTCTGTGAGTTCTGCAAG 120

Db 297 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTGGTCTGTGAGTTCTGCAAG 238

QY 121 CGCTAGTGTCTAAGCGCGTCTGTGTACAACTTAACCTGCTGCGCAAGACACAGTTGGGCC 180

Db 237 CGCTAGTGTCTAAGCGCGTCTGTGTACAACTTAACCTGCTGCGCAAGACACAGTTGGGCC 178

QY 181 TTTCGACCACTAGAACAACTTTTTCATATGACAGTTGAGAGTTGTTTAA 240

Db 177 TTTCGACCACTAGAACAACTTTTTCATATGACAGTTGAGAGTTGTTTAA 118

QY 241 CATTTGATCTTTTCTAATGAGTTAGCAGTATCTTTTTCATATGATGATTAATAATCT 300

Db 117 CATTTGATCTTTTCTAATGAGTTAGCAGTATCTTTTTCATATGATGATTAATAATCT 58

QY 301 TGAATCATATAAAAAAATAAAAAATGCTTTTGGAACTTGAATAAAAA 348

Db 57 TGAATCATATAAACTGGTTCATGCTGTGTCTTTGGAACTTGAATAAAAA 10

RESULT 2

BF554834

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

1 (Bases 1 to 493)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

cDNA Library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LLNL (info@image.llnl.gov). IMAGE ID= 1770064 The following

repetitive elements were found in this cDNA sequence: 463-485,

>AT rich#Low complexity

Seq primer: M13 Forward.

Location/Qualifiers

1. .493

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-E0-cc-c-06-0-UI"

/dev_stage="embryonic"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-E0"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: NotI; Site 2: EcoRI; This library

consists of a mixture of individually tagged normalized

libraries constructed from 8, 12 and 18-day embryo. The

tag is a string of 3-5 nucleotides present between the

Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within

the mixture."

BASE COUNT 130 a 109 c 95 g 159 t

ORIGIN

Query Match 92.9%; Score 323.4; DB 10; Length 493;

Best Local Similarity 99.7%; Pred. No. 5.1e-37;

Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAACATCACTGTCTTCCAGAGTACCAACTGTCATGTGATGATCGCGCCCTCT 60

Db 157 TGAACATCACTGTCTTCCAGAGTACCAACTGTCATGTGATGATCGCGCCCTCT 216

QY 61 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTGGTCTGTGAGTTCTGTCAG 120

Db 217 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTGGTCTGTGAGTTCTGTCAG 276

QY 121 CGCTAGTGTCTAAGCGCGTCTGTGTACAACTTAACCTGCTGCGCAAGACACAGTTGGGCC 180

Db 277 CGCTAGTGTCTAAGCGCGTCTGTGTACAACTTAACCTGCTGCGCAAGACACAGTTGGGCC 336

QY 181 TTTCGACCACTAGAACAACTTTTTCATATGACAGTTGAGAGTTGTTTAA 240

337 TTTGGACCACTAGAACCAACATTTTTCAAATGACAGTTGCAGAAATGTGGAGTGTTTTA 396 Db
241 CATTGATCTTTTGTCTAATGCAGTTAGCAGTAGTATTTTGCAGTAGTACTTAATAAATCCT 300 QY
397 CATTGATCTTTTGTCTAATGCAGTTAGCAGTAGTATTTTGCAGTAGTACTTAATAAATCCT 456 Db
301 TGAATCATATAAAAAAAAAAAAAA 325 QY
457 TGAATCATATAAAAAAAAAAAAAA 481 Db

RESULT 3	BM389059/c	BM389059	687 bp	mRNA	linear	EST 17-JAN-2002
LOCUS		UI-R-DZ0-cko-j--07-0-UI.s1	UI-R-DZ0	Rattus norvegicus	cDNA clone	
DEFINITION		UI-R-DZ0-cko-j--07-0-UI 3', mRNA sequence.				

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BM389059	BM389059.1	GI:18189112	Rattus norvegicus (Norway rat)	Rattus norvegicus (Norway rat)
	EST.		Rattus norvegicus	Rattus norvegicus
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	

REFERENCE
1 (bases 1 to 687)
AUTHORS
Bonaldi, M.F., Lennon, G. and Soares, M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)

8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: paulo-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized cartilaginous tumor library cDNA Library Preparation: M.B. Soares Lab Clone distribution: Clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

```

FEATURES
source
Location/Qualifiers
1. .687
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Bawley"
/db_xref="taxon:10116"
/clone="UI-R-D20-cko-j-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-D20"
/note="Vector: pT7T3D-pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; UI-R-D20 is a
non-normalized Rat cartilaginous tumor library (RCT)
constructed in pT7T3 PAC vector according to the procedure
described by Bonaldo Lennon & Soares (Genome Research
Genome 6: 731-806, 1996). The oligonucleotide used to
prime first strand synthesis contained the sequence tag
CAGTCCTGTA between the Not I cloning site and pT18
stretch. The Rat cartilaginous tumor tissue was provided
by Dr Jeff Stevens at the University of Iowa
TAG LIB=UI-R-D20

```

BASE COUNT	222 a	130 c	163 g	172 t
TAG_SEQ=CATCTTGTA"				
TAG_TISSUE=cartilaginous tumor				
ARG_EID=OI-K-DZU				

[illegible]

	BQ780699	linear	EST 26-JUL-2002
RESULT 4	UI-R-FFO-cpc-k-22-0-UI.s1	UT-R-FFO Rattus norvegicus cDNA clone	
BQ780699/c	UI-R-FFO-cpc-k-22-0-UI 3'	mRNA sequence.	
LOCUS	459 bp		
DEFINITION	Rattus norvegicus cDNA clone		

ACCESSION	BQ780699
VERSION	BQ780699.1
KEYWORDS	GI:21989171
SOURCE	EST.
ORGANISM	Rattus norvegicus (Norway rat)
	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
REFERENCE	1 (bases 1 to 459)
AUTHORS	Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477
 PUBMED 8889548
 COMMENT

Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu

Tissue Procurement: Jeff Stevens
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: DISTRIBUTION: Researchers may obtain clones
 from Research Genetics (www.resgen.com).
 Seg primer: M13 FORWARD

```

FEATURES
source
POLYRES.
Location/Qualifiers
1..459
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="U1-R-PF0-cpc-k-22-0-UI"
/tissue_type="Mixed tissues"

```

/dev stage="Adult"
/lab host="DH10B (Life Technologies) (T1 phage resistant)"
/clone lib="UI-R-PF0"
/note="Vector: p77T3-Pac (Pharmacia) with a modified
polylinker; Site 1: Ecor I; Site 2: Not I; UI-R-PF0 is a
subtracted cDNA library containing the following tissue(s)
) : Normal cartilage and SR-JWS Tumor line . The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTAATGGAGC,
CATCTTCTGA.
TAG LIB=UI-R-PF0
TAG TISSUE=cartilage
TAG_SEQ=CTAATGGAGC"
BASE COUNT 145 a 91 c 103 g 120 t
ORIGIN

Query Match 92.5%; Score 321.8; DB 13; Length 459;
Best Local Similarity 99.4%; Pred. No. 9e-37;
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 60
Db 325 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 266
QY 61 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 120
Db 265 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 206
QY 121 CGCTAGTGTCTAACGGCGTCTGTACAACTCACTCACTGGCAAGACACAGTGTGGGCC 180
Db 205 CGCTAGTGTCTAACGGCGTCTGTACAACTCACTCACTGGCAAGACACAGTGTGGGCC 146
QY 181 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAAGTGTGGAGTGTTTTA 240
Db 145 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAAGTGTGGAGTGTTTTA 86
QY 241 CATTGATCTTTGCTAATGCAGTGTAGCAGTATGTTTTCATGATGATGATGATGATGAT 300
Db 85 CATTGATCTTTGCTAATGCAGTGTAGCAGTATGTTTTCATGATGATGATGATGATGAT 26
QY 301 TGAATCATATAAAAAAAAAAAAAA 325
Db 25 TGAATCATATAAAAAAAAAAAAAA 1

RESULT 5
BQ780117/c
LOCUS
DEFINITION
UI-R-PF0-cow-j-10-0-UI.sl mRNA linear EST 26-JUL-2002
UI-R-PF0-cow-j-10-0-UI.sl UI-R-PF0 Rattus norvegicus cDNA clone
BQ780117
VERSION
BQ780117.1 GI:21988589
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 640)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS
Normalization and subtraction: two approaches to facilitate gene
discovery
TITLE
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=Yes.

Location/Qualifiers
1. .640
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="UI-R-PF0-cow-j-10-0-UI"
/tissue="Type=Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-PF0"
/note="Vector: p77T3-Pac (Pharmacia) with a modified
polylinker; Site 1: Ecor I; Site 2: Not I; UI-R-PF0 is a
subtracted cDNA library containing the following tissue(s)
) : Normal cartilage and SR-JWS Tumor line . The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTAATGGAGC,
CATCTTCTGA.
TAG LIB=UI-R-PF0
TAG TISSUE=rat SRC-JWS tumor line
TAG_SEQ=CATCTTCTGA"
BASE COUNT 208 a 121 c 152 g 159 t
ORIGIN

Query Match 92.5%; Score 321.8; DB 13; Length 640;
Best Local Similarity 99.4%; Pred. No. 7.5e-37;
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 60
Db 325 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 267
QY 61 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 120
Db 266 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 207
QY 121 CGCTAGTGTCTAACGGCGTCTGTACAACTCACTCACTGGCAAGACACAGTGTGGGCC 180
Db 206 CGCTAGTGTCTAACGGCGTCTGTACAACTCACTCACTGGCAAGACACAGTGTGGGCC 147
QY 181 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAAGTGTGGAGTGTTTTA 240
Db 146 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAAGTGTGGAGTGTTTTA 87
QY 241 CATTGATCTTTGCTAATGCAGTGTAGCAGTATGTTTTCATGATGATGATGATGATGAT 300
Db 86 CATTGATCTTTGCTAATGCAGTGTAGCAGTATGTTTTCATGATGATGATGATGATGAT 27
QY 301 TGAATCATATAAAAAAAAAAAAAA 325
Db 26 TGAATCATATAAAAAAAAAAAAAA 2

RESULT 6
AWS20555/c
LOCUS
DEFINITION
UI-R-BJOp-afx-g-09-0-UI.sl mRNA linear EST 06-MAR-2000
UI-R-BJOp-afx-g-09-0-UI.sl UI-R-BJOp Rattus norvegicus cDNA clone
UI-R-BJOp-afx-g-09-0-UI.3', mRNA sequence.
ACCESSION
AWS20555

```

VERSION      AW520555.1  GI:7162933
SOURCE       Rattus norvegicus (Norway rat)
ORGANISM     Rattus norvegicus
REFERENCE    1 (bases 1 to 477)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED      8889548
COMMENT      Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              The sequence contained an oligo-dT track that was present in the
              oligonucleotide that was used to prime the synthesis of first
              strand cDNA and therefore this may represent a bonafide poly A
              tail. The sequence tag present in the cDNA between the NotI site
              and the oligo-dT track served to identify it as a clone from the
              normalized AV canal at 15 dpc library cDNA Library Preparation:
              M.B. Soares Lab Clone distribution: clones will be available
              through Research Genetics (www.resgen.com)
              Seq primer: M13 Forward
              POLYA=Yes.

FEATURES     Location/Qualifiers
             1..477
             /organism="Rattus norvegicus"
             /mol_type="mRNA"
             /strain="Sprague-Dawley"
             /db_xref="taxon:10116"
             /clone="UI-R-BJ0p-afx-g-09-0-UI"
             /dev_stage="adult"
             /lab_host="DH10B (Life Technologies)"
             /clone_lib="UI-R-BJ0p"
             /note="Vector: pT73D-Pac (Pharmacia) with a modified
             polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0p
             library is a subtracted library derived from the UI-R-AA1,
             UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
             UI-R-AG1 libraries. These libraries represent tissues from
             rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
             at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
             canal at 15 dpc, and ventricle at 13 dpc. The tag is a
             string of 5-6 nucleotides present between the Not I site
             and the oligo-dT track. The library was constructed as
             described by Bonaldo, Lennon and Soares, Genome Research
             6:791-806, 1996.
             TAG_LIB=UI-R-BJ0p
             TAG_TISSUE=AV canal at 15 dpc
             TAG_SEQ=CAAGG"

BASE COUNT   157 a 103 g 124 t
ORIGIN
Query Match      92.4%; Score 321.4; DB 9; Length 477;
Best Local Similarity 99.7%; Pred. No. 1e-36;
Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  TGAACATCACTGCTTCCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCT 60
DB      323 TGAACATCACTGCTTCCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCT 264
QY      61  AGACCTACCCACGGGACACATGCTTCGGTACCTTTGGGCTCTGAGGTTCTGTCAG 120
DB      263  AGACCTACCCACGGGACACATGCTTCGGTACCTTTGGGCTCTGAGGTTCTGTCAG 204
QY      121 CGTAGTCTAACGCCGTTCTGTACAACTCACTGCGCAAGAACACAGTGTGGGCC 180

```

```

Db      203 CGTAGTCTAACGCCGTTCTGTACAACTCACTGCGCAAGAACACAGTGTGGGCC 144
QY      181 TTTCGACCACTAGAACAACTTTTTCATTCACAGTTGCAGAAATGTGGAGTGTTTTA 240
DB      143 TTTCGACCACTAGAACAACTTTTTCATTCACAGTTGCAGAAATGTGGAGTGTTTTA 84
QY      241 CATTGATCTTTTGTCTAATGTCAGTATGATGATGATGATGATGATGATGATGATGAT 300
DB      83 CATTGATCTTTTGTCTAATGTCAGTATGATGATGATGATGATGATGATGATGATGAT 24
QY      301 TGAATCATATAAAAAA 323
DB      23 TGAATCATATAAAAAA 1

RESULT 7
AA859879/c
LOCUS      AA859879
DEFINITION Rattus norvegicus (Norway rat) linear EST 03-JUL-1999
            UI-R-E0-cc-c-06-0-UI.51 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-cc-c-06-0-UI 3', similar to dbj|D25274|HUMPO2ST9 Human
            randomly sequenced mRNA, mRNA sequence.
ACCESSION  AA859879
VERSION     AA859879.1 GI:4230416
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 489)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949399.
            Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics This clone is also available through the
            I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE
            ID=1770064
            Seq primer: M13 Forward
            POLYA=No.

FEATURES     Location/Qualifiers
             1..489
             /organism="Rattus norvegicus"
             /mol_type="mRNA"
             /strain="Sprague-Dawley"
             /db_xref="taxon:10116"
             /clone="UI-R-E0-cc-c-06-0-UI"
             /dev_stage="embryonic"
             /lab_host="DH10B (Life Technologies)"
             /clone_lib="UI-R-E0"
             /note="Vector: pT73D-Pac (Pharmacia) with a modified
             polylinker; Site 1: NotI; Site 2: EcoRI; This library
             consists of a mixture of individually tagged normalized
             libraries constructed from 8, 12 and 18-day embryo. The
             tag is a string of 3-5 nucleotides present between the
             Not I site and the oligo-dT track which allows
             identification of the library of origin of a clone within
             the mixture."

BASE COUNT   158 a 96 c 107 g 128 t
ORIGIN

```


Tel: 319 335 8250
Fax: 319 335 9565

Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized osteoblast library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source 1..760
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DRI-cky-h-11-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-DRI"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DRI library is a normalized Rat Osteoblast library (nREO) constructed in pT377 vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AAGATACAA between the Not I cloning site and dT18 stretch. The Rat Osteoblast tissue was provided by Lian & Stein of the University of Massachusetts Medical School.
TAG LIB=UI-R-DRI
TAG TISSUE=osteoblast
TAG SEQ=AAGATACAA"
BASE COUNT 244 a 173 g 198 t 2 others
ORIGIN

Query Match 92.4%; Score 321.4; DB 13; Length 760;
Best Local Similarity 99.4%; Pred. No. 7.7e-37;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGAACATCACTCTCTGCGAGTACCACTGTCATGTGATGCGGCCCTCT 60
Db |||||||
324 TGAACATCACTCTCTGCGAGTACCACTGTCATGTGATGCGGCCCTCT 265
QY 61 AGACCTCACCACGCGGACACATGCTTCGGTACCTTTGGTCTGTGAGTTCTGTCAAG 120
Db |||||||
264 AGACCTCACCACGCGGACACATGCTTCGGTACCTTTGGTCTGTGAGTTCTGTCAAG 205
QY 121 CGCTAGTGTACGCGGTTCTGTACACCTTACTCACTGCGAAGACACAGTTGGGCC 180
Db |||||||
204 CGCTAGTGTACGCGGTTCTGTACACCTTACTCACTGCGAAGACACAGTTGGGCC 145
QY 181 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTGCGAATTTGTGAGTGTTTTA 240
Db |||||||
144 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTGCGAATTTGTGAGTGTTTTA 85
QY 241 CATTTGATCTTTTCTAATGAGTTAGCAGTATCTTTTTCATGATGATCACTTAATAATCT 300
Db |||||||
84 CATTTGATCTTTTCTAATGAGTTAGCAGTATCTTTTTCATGATGATCACTTAATAATCT 25
QY 301 TGAATCATAAAAAATAAAAAA 324
Db |||||||
24 TGAATCATAAAAAATAAAAAA 1

RESULT 10
BQ190077/c
LOCUS

BQ190077 472 bp mRNA linear EST 30-APR-2002

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

UI-R-CNI-cjs-h-07-0-UI-s3 UI-R-CNI Rattus norvegicus cDNA clone
UI-R-CNI-cjs-h-07-0-UI 3', mRNA sequence.
BQ190077
BQ190077.1 GI:20365628
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 472)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat eye library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers
1..472
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CNI-cjs-h-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CNI"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CNI library is a subtracted library derived from the following pool of seven normalized rat libraries: normalized rat seminal vesicles, normalized rat penis, normalized rat bladder, normalized rat fundus, normalized rat brown adipose, normalized rat cervix, and normalized rat salivary gland. It was constructed according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the CNI library, plasmid DNA from the pool of normalized libraries was electroporated into competent bacteria for the production of single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat Unigene Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and MN. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subtracted libraries CA0 and CA1 corresponding to plates R-CA0-ANV through R-CA0-AXS, R-CA0-AZX through R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS, R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA, R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through R-CA0-BOJ, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through R-CA1-BDA, R-CA1-BHZ through R-CA1-BJF, R-CA1-BJR, R-CA1-BUT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF, R-CA1-BKI, R-CA1-BKT, R-CA1-BLF, R-CA1-BLH through

R-CA1-BLN, R-CA1-BLS, R-CA1-BLU-V, R-CA1-BNR, and R-CA1-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries CS0s, CT0s, CU0s, CW0s, CX0s and normalized libraries CS0, CT0, CU0, CW0, and CX0 corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN, R-CU0-BUQ through R-CU0-BVL, R-CW0-BVV through R-CW0-BWP, R-CX0-BXN through R-CW0-BXO, R-CX0-BWQ through R-CX0-BXM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library CV0 and 4,000 from normalized eye library CV1) corresponding to plates R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BV0 and BVOp (7-9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BVOp-AOI through R-BVOp-AOX, and R-BX0-AQY through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CV0, CZ0, DA0, DB0, DC0, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ through R-DB0-BZA, R-DC0-BZI through R-DC0-BZQ, R-DC0-CAY through R-DC0-CBA, R-DD0-BZR through R-DD0-CAA, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CN0, that makes up the tracer. The corresponding plates are R-CN0-BKW through R-CN0-BLD, R-CN0-BLG, R-CN0-BLP through R-CN0-BLR, R-CN0-BLT, R-CN0-BMW-X, R-CN0-BMB, and R-CN0-BMF through R-CN0-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CN0 pool corresponding to the following addresses: bkx-a-09-0-UI, bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI, bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI, bky-a-05-0-UI, bkz-a-06-0-UI, bkz-a-11-0-UI, bkz-c-06-0-UI, bkz-c-09-0-UI, bkz-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI, blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CN0 clone (corresponding to the address bkz-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.

TAG LIB=UI-R-CN1
TAG TISSUE=rat eye
TAG SEQ=CAGCC

BASE COUNT 154 a 102 g 123 t
ORIGIN

Query Match 92.2%; Score 320.8; DB 13; Length 472;
Best Local Similarity 99.4%; Pred. No. 1.2e-36;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAACATCACTGTTGGCAGAGTACCAACACTGTCATGTCATTGATGCGGCCCTCT 60
Db 324 TGAACATCACTGTTGGCAGAGTACCAACACTGTCATGTCATTGATGCGGCCCTCT 265
QY 61 AGACCTCACCCACGGGACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 120

Db 264 AGACCTCACCCACGGGACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 205
QY 121 CGCTAGTCTTAAGCGCGTTCTGTCAACCTAACTCACTGGCAAGAACACAGTGTGGGCC 180
Db 204 CGTAGTCTTAAGCGCGTTCTGTCAACCTAACTCACTGGCAAGAACACAGTGTGGGCC 145
QY 181 TTTCGACCACCTAGAACAAACTTTTTCATTTGACAGTGTGCAGAAATGTGGAGTGTTTTA 240
Db 144 TTTCGACCACCTAGAACAAACTTTTTCATTTGACAGTGTGCAGAAATGTGGAGTGTTTTA 85
QY 241 CATTGATCTTTGCTTAATGCGAGTGTAGCAGTATGTTTCGATGACTTAATAATCT 300
Db 84 CATTGATCTTTGCTTAATGCGAGTGTAGCAGTATGTTTCGATGACTTAATAATCT 25
QY 301 TGAATCATATAAAAAA 324
Db 24 TGAATCATATAAAAAA 1
RESULT 11
BG380448/c
LOCUS
DEFINITION
UI-R-CT0-btx-d-05-0-UI.s1 UI-R-CT0 Rattus norvegicus cDNA clone
UI-R-CT0-btx-d-05-0-UI 3', mRNA sequence.
ACCESSION
BG380448
VERSION
BG380448.1 GI:13304920
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 615)
Bonaldo, M.P., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
97044477
MEDLINE
8889548
PUBMED
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized rat brain pool library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..615
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CT0-btx-d-05-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CT0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CT0
library is a normalized library constructed from the
following rat brain tissues: embryonic day 17, embryonic
day 19, embryonic day 21, adult day 1, adult day 12, adult
day 75, adult day 200. For a detailed description of the

library from which this clone was derived, please visit our web site at rateg.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG LIB=UI-R-CT0
TAG_TISSUE=rat brain pool
TAG_SEQ=ACTTC

BASE COUNT 200 a 117 c 139 g 158 t 1 others
ORIGIN

Query Match 92.2%; Score 320.8; DB 10; Length 615;
Best Local Similarity 99.1%; Pred. No. 1.1e-36;
Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTCCAGAGTACCAACACTGCTATGATGATGCGGCCCTCT 60
Db 326 TGAACATCACTGCTTCCAGAGTACCAACACTGATGATGATGCGGCCCTCT 267
QY 61 AGACTCACCCACCGGACACATGCTTCCGTTACCTTTGGGTCTGTGAGGTTCTGTCAG 120
Db 266 AGACTCACCCACCGGACACATGCTTCCGTTACCTTTGGGTCTGTGAGGTTCTGTCAG 207
QY 121 CGCTAGTCTTAACCGCTTCTGTACACCTTAACCTACCTGCGCAAGACACAGTGTGGGCC 180
Db 206 CGCTAGTCTTAACCGCTTCTGTACACCTTAACCTACCTGCGCAAGACACAGTGTGGGCC 147
QY 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCAGATTTGAGAGTGTGTTTA 240
Db 146 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCAGATTTGAGAGTGTGTTTA 87
QY 241 CATTGATCTTTTGTCTAATGAGTGTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300
Db 86 CATTGATCTTTTGTCTAATGAGTGTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 27
QY 301 TGAATCATRAAAAAAAAAAAAAAAAAA 325
Db 26 TGAATCATRAAAAAAAAAAAAAAAAAA 2

RESULT 12
BE111691/c

LOCUS BE111691 639 bp mRNA linear EST 13-JUN-2000
DEFINITION UI-R-BJ1-avv-f-10-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
UI-R-BJ1-avv-f-10-0-UI 3', mRNA sequence.

ACCESSION BE111691
VERSION BE111691.1 GI:8503796
KEYWORDS EST
SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 639)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477

JOURNAL MEDLINE
PUBMED

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized ventricle at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available

through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..639
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-avv-f-10-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ1"

/note=vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ1 library is a subtracted library derived from the following tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, ventricle at 13 dpc, and adult heart. For a detailed description of the library from which this clone was derived, please visit our web site at rateg.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG LIB=UI-R-BJ1
TAG_TISSUE=ventricle at 13 dpc
TAG_SEQ=CAGCGA

BASE COUNT 209 a 121 c 151 g 157 t 1 others
ORIGIN

Query Match 92.2%; Score 320.8; DB 10; Length 639;
Best Local Similarity 99.1%; Pred. No. 1e-36;
Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTCCAGAGTACCAACACTGATGATGATGCGGCCCTCT 60
Db 325 TGAACATCACTGCTTCCAGAGTACCAACACTGATGATGATGCGGCCCTCT 265
QY 61 AGACTCACCCACCGGACACATGCTTCCGTTACCTTTGGGTCTGTGAGGTTCTGTCAG 120
Db 265 AGACTCACCCACCGGACACATGCTTCCGTTACCTTTGGGTCTGTGAGGTTCTGTCAG 206
QY 121 CGCTAGTCTTAACCGCTTCTGTACACCTTAACCTACCTGCGCAAGACACAGTGTGGGCC 180
Db 205 CGCTAGTCTTAACCGCTTCTGTACACCTTAACCTACCTGCGCAAGACACAGTGTGGGCC 146
QY 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCAGATTTGAGAGTGTGTTTA 240
Db 145 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCAGATTTGAGAGTGTGTTTA 86
QY 241 CATTGATCTTTTGTCTAATGAGTGTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300
Db 85 CATTGATCTTTTGTCTAATGAGTGTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 26
QY 301 TGAATCATRAAAAAAAAAAAAAAAAAA 325
Db 25 TGAATCATRAAAAAAAAAAAAAAAAAA 1

RESULT 13
BQ200122/c

LOCUS BQ200122

DEFINITION UI-R-EB1-clf-i-07-0-UI.s1 UI-R-EB1 Rattus norvegicus cDNA clone

UI-R-EB1-clf-i-07-0-UI 3', mRNA sequence.

ACCESSION BQ200122

VERSION BQ200122.1 GI:20416587

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 705)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

BQ200122 705 bp mRNA linear EST 02-MAY-2002
UI-R-EB1-clf-i-07-0-UI.s1 UI-R-EB1 Rattus norvegicus cDNA clone
UI-R-EB1-clf-i-07-0-UI 3', mRNA sequence.

BQ200122.1 GI:20416587
EST.

Rattus norvegicus (Norway rat)
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 705)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized duodenum library cDNA Library Preparation: M.B. Soares Lab Clone distribution: Clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..705
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-EB1-clf-i-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-EB1"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; UI-R-EB1 is a non-normalized Rat Duodenum library (RDU) constructed in pT73 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag GTGGTTTCAT between the Not I cloning site and dT18 stretch. The Rat Duodenum tissue was provided by Tom Freeman of the Sanger Center.
TAG_LIB=UI-R-EB1
TAG_TISSUE=duodenum
TAG_SEQ=GTGGTTTCAT
BASE COUNT 226 a 136 c 165 g 177 t 1 others
ORIGIN

Query Match 92.2%; Score 320.8; DB 13; Length 705;
Best Local Similarity 99.1%; Pred. No. 9.8e-37;
Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGAACATCACTCTCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCT 60
Db 326 TGAACATCACTCTCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCT 267
QY 61 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 120
Db 266 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 207
QY 121 CGCTAGTGTAAACCGCGTCTGTACACCTAACTCACTGGCAAGACACAGTGTGGGCC 180
Db 206 CGCTAGTGTAAACCGCGTCTGTACACCTAACTCACTGGCAAGACACAGTGTGGGCC 147
QY 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGAGTTGAGTGTGTTT 240
Db 146 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGAGTTGAGTGTGTTT 87
QY 241 CATTTGATCTTTTGCATATGAGTTAGCAGTATGTTTTCATGATGATGATTAATTCCT 300
Db 86 CATTTGATCTTTTGCATATGAGTTAGCAGTATGTTTTCATGATGATGATTAATTCCT 27
QY 301 TGAATCATATAAAAAA 325

Db 26 TGAATCATATAAAAAA 2
RESULT 14
AA899757/c
LOCUS
DEFINITION
UI-R-E0-da-b-11-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA Clone
UI-R-E0-da-b-11-0-UI 3', similar to gi|464185|dbj|D25274|HUMPO2ST9
Homc sapiens mRNA, clone: P02ST9, mRNA sequence.
ACCESSION
AA899757
VERSION
AA899757.1 GI:4232251
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (Bases 1 to 465)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
COMMENT
On Apr 7, 1998 this sequence version replaced gi:3035111.
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
CDNA Library Preparation: M. Ratima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward
Location/Qualifiers
1..465
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-da-b-11-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-E0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
BASE COUNT 155 a 93 c 99 g 118 t
ORIGIN

Query Match 91.4%; Score 318.2; DB 9; Length 465;
Best Local Similarity 99.1%; Pred. No. 2.9e-36;
Matches 320; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGAACATCACTCTCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCT 60
Db 323 TGAACATCACTCTCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCT 264
QY 61 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 120
Db 263 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 204
QY 121 CGCTAGTGTAAACCGCGTCTGTACACCTAACTCACTGGCAAGACACAGTGTGGGCC 180
Db 203 CGCTAGTGTAAACCGCGTCTGTACACCTAACTCACTGGCAAGACACAGTGTGGGCC 144

QY 181 TTTCGACCACCTAGAACAACTTTTCAATTGACAGTTGCAGAAATGTGGAGTGTTTTA 240
 Db 143 TTTCGACCACCTAGAACAACTTTTCAATTGACAGTTGCAGAAATGTGGAGTGTTTTA 84
 QY 241 CAATTGATCTTTTGTGTAATGCACTAGCAGTATGTTTTCATGTATGACTTAATAAATCCT 300
 Db 83 CAATTGATCTTTTGTGTAATGCACTAGCAGTATGTTTTCATGTATGACTTAATAAATCCT 24
 QY 301 TGAATCATATAAAAAA 323
 Db 23 TGAATCATATAAAAAA 1

RESULT 15

BE104696/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BE104696 405 bp mRNA linear EST 13-JUN-2000
 UI-R-BX0-ars-e-01-0-UI.s1 UI-R-BX0 Rattus norvegicus cDNA clone
 UI-R-BX0-ars-e-01-0-UI 3', mRNA sequence.

BE104696

BE104696.1 GI:8496796

EST.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa;

Mammalia; Eutheria;

Chordata; Craniata;

Vertebrata; Euteleostomi;

Mammalia; Eutheria;

Rodentia; Sciurognathi;

Muridae; Murinae;

Rattus.

1 (bases 1 to 405)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

The following repetitive elements were found in this cDNA sequence:

1-22, >Ar-rich#low complexity

Seq primer: M13 Forward

POLYA=Yes.

Location/Qualifiers

1. 405

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-BX0-ars-e-01-0-UI"

/dev_stage="embryonic 13 dpc"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-BX0"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BX0

library is derived from 13 dpc whole embryo tissue. For a

detailed description of the library from which this clone

was derived, please visit our web site at

ratseq.eng.uiowa.edu.

TAG_SEQ=None found"

BASE COUNT 131 a 87 g 110 t 1 others

ORIGIN

Query Match

Best Local Similarity 91.1%; Score 317; DB 10; Length 405;

Matches 320; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGAACATCACTGTCTTGGCCAGAGTACCAACACTGTCAATGTGATGATCGCCGCCCTCT 60
 Db 328 TGAACATCACTGTCTTGGCCAGAGTACCAACACTGTCAATGTGATGATCGCCGCCCTCT 269
 QY 61 AGACCTCACCCACCGGACACATGCTTCCGGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 120
 Db 268 AGACCTCACCCACCGGACACATGCTTCCGGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 209
 QY 121 CGCTAGTGTAAACCGCGTCTGTGTACAACTCACTGGCAAGAACACAGTGTGGGCC 180
 Db 208 CGCTAGTGTAAACCGCGTCTGTGTACAACTCACTGGCAAGAACACAGTGTGGGCC 149
 QY 181 TTTCGACCACCTAGAACAACTTTTCAATTGACAGTTGCAGAAATGTGGAGTGTTTTA 240
 Db 148 TTTCGACCACCTAGAACAACTTTTCAATTGACAGTTGCAGAAATGTGGAGTGTTTTA 89
 QY 241 CAATTGATCTTTTGTGTAATGCACTAGCAGTATGTTTTCATGTATGACTTAATAAATCCT 300
 Db 88 CAATTGATCTTTTGTGTAATGCACTAGCAGTATGTTTTCATGTATGACTTAATAAATCCT 29
 QY 301 TGAATCATATAAAAAA 325
 Db 28 TGAATCTTTATATAAAAAA 4

Search completed: November 23, 2003, 15:50:08

Job time : 820.065 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:43:52 ; Search time 2390.73 Seconds
(without alignments)
10338.975 Million cell updates/sec.

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccattttgttcagatt.....cccttggtctgtgaggttc 1017

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_estc:**

9: gb_esti:**

10: gb_est2:**

11: gb_est3:**

12: gb_est4:**

13: gb_est5:**

14: gb_est6:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_mam:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_pbg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678.2	66.7	1796	11 AK081613	AK081613 Mus muscu
2	677.6	66.6	2269	11 AK034601	AK034601 Mus muscu
3	677.6	66.6	2270	11 AK047969	AK047969 Mus muscu
4	660.2	64.9	2308	11 AK076023	AK076023 Mus muscu

5	594.2	58.4	978	13	BUS14423	BUS14423
6	562	55.3	826	14	CA318577	CA318577
7	558.2	54.9	751	13	BQ177713	BQ177713
8	552.2	54.3	960	13	BUS11766	BUS11766
9	551.2	54.2	557	10	BG665146	BG665146
10	537.4	52.8	793	14	CA319273	CA319273
11	535.2	52.6	897	12	B1646446	B1646446
12	528.6	52.0	714	12	B1851489	B1851489
13	528.6	52.0	915	13	BQ931374	BQ931374
C 14	526	51.7	545	9	AA817948	AA817948
C 15	526	51.7	639	13	BQ201343	BQ201343
C 16	525.4	51.7	760	13	BQ191985	BQ191985
C 17	510.4	50.2	806	13	BQ444803	BQ444803
18	508.2	50.0	515	14	CD372599	CD372599
19	499.4	49.1	644	14	CA339407	CA339407
20	493.8	48.6	686	14	CB057853	CB057853
21	490.8	48.3	672	12	B1692509	B1692509
22	485	47.7	969	10	BF783603	BF783603
23	483.4	47.5	680	14	CD353147	CD353147
24	482.8	47.5	741	12	B1694684	B1694684
25	482.4	47.4	732	14	CB596051	CB596051
C 26	479	47.1	502	9	AI028951	AI028951
27	474.8	46.7	787	12	BG969782	BG969782
C 28	473.4	46.5	705	13	BQ200122	BQ200122
C 29	472.4	46.5	687	12	BM389059	BM389059
30	462.4	45.5	796	12	B1735385	B1735385
31	458.4	45.1	588	14	CA553516	CA553516
32	454.2	44.7	805	12	B1692090	B1692090
33	453	44.5	594	14	CB055806	CB055806
34	450.8	44.3	667	12	BQ042534	BQ042534
35	449.2	44.2	938	13	BQ228330	BQ228330
36	442.8	43.5	447	10	BG663683	BG663683
37	438.8	43.1	671	12	BM946306	BM946306
38	438.2	43.1	704	12	BM951935	BM951935
39	427.2	42.0	538	12	BM053597	BM053597
40	426.6	41.9	550	12	BM022481	BM022481
41	426.2	41.9	950	10	BF168374	BF168374
42	426	41.9	556	12	B1655954	B1655954
C 43	425.4	41.8	575	10	BF659623	BF659623
C 44	423.8	41.7	640	13	BQ780117	BQ780117
C 45	422	41.5	702	13	BQ176893	BQ176893

ALIGNMENTS

RESULT 1
AK081613
LOCUS
DEFINITION
AK081613
Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:Cl30051H24 product:RAS-related C3 botulinum substrate 1, full insert sequence.
ACCESSION
AK081613
VERSION
AK081613.1 GI:26349264
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Muridae; Sciurognathi; Muridae; Mus.
REFERENCE
1
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
2049374
PUBMED
11042159


```

Db      1146 CATCCCTTGGCGCCGCCCCACCTCCACTCCGCTCCCTCCGCCAGATGGGCGCTTGAGTCCA 1205
QY      643 TATTAACTGGCCATCCTCACAGTCTGCTACTTACCTAGCAAGTCTTTCTTTAGACACCCC 702
Db      1206 TATTAACTGGCCACCCCTCACAGTT-CTAACTTAGCAAGTCTTTCTTTA-GAACCCCC 1263
QY      703 TTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCGATAATGCA-----T 756
Db      1264 TTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCGATAACCGCAGAGATT 1323
QY      757 TCGAGAGTTTTTTGGTAGATAGTAGAAGTGGTTCCTGTTTTCCACCTTCTTTACTCAG 816
Db      1324 TTGTGTTTTTTTTGGTAGTCTCAGTAGAAGTGGTCTGTTTTTCGCTTACTTTACTCAG 1383
QY      817 -CTGACTAGTGTCTCCCTTCGTTTTCTAGTAACTGGGTGAGTAATCAGTGTGGCGCT 875
Db      1384 AGCTAGTGTGCTTCTTCTAGTTTCCCTAGCAACTAGGTGTGCAAAATCATGTGTTCGACGT 1443
QY      876 TTACAGTTTTTAACTATTTTAGATATTTCTGAACATCACTGTCTTGCAGAGTACCAAC 935
Db      1444 TTACGGTTTTTAACTATTTTAGATATTTCTAACTATGAACTTCTAAACAGTACCAAC 1503
QY      936 ACTGTCATGTATGATGATGCGCCGCCCTCTAGACCTACCCACGCGACACATGCTTCCGG 995
Db      1504 ACTGTCATGTGACTAATGCTG---CCTCTAGACCTCGCCACGTCGACAGCGCTTCCG 1560
QY      996 T 996
Db      1561 T 1561

RESULT 2
AC0334601
LOCUS      2269 bp      mRNA      linear      HTC 05-DEC-2002
DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region
            and neck cDNA, RIKEN full-length enriched library, clone:9430013G17
            product:RAS-related C3 botulinum substrate 1, full insert sequence.
ACCESSION  AK034601
VERSION     AK034601.1 GI:26330056
KEYWORDS   HTC; CAP trapper.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1
AUTHORS    Carninci, P. and Hayashizaki, Y.
TITLE      High-efficiency full-length cDNA cloning
JOURNAL    Meth. Enzymol. 303, 19-44 (1999)
MEDLINE    99279253
PUBMED     10349636
AUTHORS    Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE      Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE    20493374
PUBMED     11042159
AUTHORS    Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
            Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
            Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
            Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
            Fujiwaki, S., Inoue, K., Iogawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
            Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE      RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE    20530913
PUBMED     11076861
AUTHORS    Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,

```

```

Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, K., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
King, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2269)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. .2269
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:9430013G17"
/db_xref="taxon:10090"
/clone="9430013G17"
/tissue_type="embryonic body between diaphragm region and
neck"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
198. .776
/note="unnamed protein product; RAS-related C3 botulinum
substrate 1 (MGI:97845, GB|AK011072, evidence: BLASTN,
100%, match=98)"

```

FEATURES
source

CDS

putative"
/codon_start=1
/protein_id="BAC28767.1"
/db_xref="GI:26330057"
/translation="MQAIKCVVGVGAVGKTCLLISYTTNAPPGEYIPIVFDNYSANV
MVDKPNLIGLMDTAGEDYDRLRLPSVQTDVFLICSLVSPASFEVRAKMYPEVR
HHCENPIILVGTIKDLDRDDKDTIEKLEKILPTIYFQGLAMAKEIGAVKYLECSAL
TORGLKTVDFDEAIRAVLCPPPPVKKRKRKCLLL"
2252..2257
/note="putative"
polyA_signal
2269
/note="putative"
polyA_site
2269
/note="putative"
BASE COUNT 555 a 569 c 502 g 643 t
ORIGIN

Query Match 66.6%; Score 677.6; DB 11; Length 2269;
Best Local Similarity 85.3%; Pred. No. 2.9e-183;
Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;
QY 1 CCCATTCTTCTGCTCAGATTAGAAATGCCAAATACCTTGTGAACCTAAGTTGCGTTGTG 60
DB 1020 CCCCATTCTTCTGCTCAGATTAGAAATGCCAAATACCTTGTGAACCTAAGTTGCGTTGTG 1079
QY 61 CTGAGAACACCTAAGCACTAAGCTGTGAGAGA---CTTGTCCCTTAAGAACACTGCAGC 117
DB 1080 CTGAGAACACCTAAGCACTAAGCTGTGAGAGACTTCTGTGAGAGACTTCTGTGTAAGAACCCGACG 1139
QY 118 TTCTGGGCTCAGGGTGCAGACCCCTCCGCTAGCTCCAGACCGTGTGACACGACAGCC 177
DB 1140 TTCTGGAGCCAGGGTGCAGACCCCTCCAGAGTTCAGACCGTGTGACACGACAGCC 1199
QY 178 TCCTTAATGACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATACGT 237
DB 1200 TCCTTAATGACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATACGT 1259
QY 238 AACTTTGACTGACTGACAC---GATGGGTGTAAACAGCTCTGCTCTTTCATTTCTAGTG 294
DB 1260 AACTTTGACTGACTGACACGATGAGTGAGTGACAGCTCAGCTCTTGAATTTCTAGTG 1319
QY 295 AGTTCTCTAAATATACAGCTGACCGCTCTCTGAGGCTTTGACAGAACTCTGGCTCCTG 354
DB 1320 AGTTCTCTAAATATACAGCTGACCGCTCTCTGAGAACTTTGACAGAACTCTGGTCTCTG 1379
QY 355 TGTGCTCTTAAGAAAGTATCTGTTCTAGTCTGGGTGTGCTGGGTGGAGTGTGTGA 414
DB 1380 TGTGCTCTTAAGAAAGTATCTGTTCTAGTGTGGGTGTGCTGGGTGGA---GTGTGA 1437
QY 415 ACACGACGTCATCAAGGAGACAGACAGTATTTTGAATATGAATGAGATTAATTT 474
DB 1438 ACACGACATGATCAAGGAGACAGACAGTATTTTGAATATGAATGAGATTAATTT 1497
QY 475 ACACATCATGTACATGAGTAAATCAACTGAATAAAGTGTACCGGTAAAGCTTTTGA 534
DB 1498 ACACATCATGTACATGAGTAAATCAACTGAATAAAGTGTACCGGTAAAGCTTTTGA 1557
QY 535 AGGTTAAATTTCTGTCAAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC 589
DB 1558 AGGTTAAATTTCTGTCAAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTTGTAG 1617
QY 590 -----TCTTGAGCCCCCTTCCCTGCTGTCCTCCCTCCAGATGGGGGCTTGAAGTCCA 642
DB 1618 CATCCCTTGGCGCCCCCACCCTCCAGTCCGCTCCCTCCAGATGGGGGCTTGAAGTCCA 1677
QY 643 TATTAAATGGCCATCTCTCAGAGTGTGCTAAGTAACTAGAGAGTGTCTTTCTTTAGGACCC 702
DB 1678 TATTAAATGGCCATCTCTCAGAGTGTGCTAAGTAACTAGAGAGTGTCTTTCTTTA-GAAGCC 1735
QY 703 TTCTTAACGACGATATGCTGACCTGTGCTAATAGATCTTCTGATATGCA-----TT 757
DB 1736 TTCTTAACGACGATATGCTGACCTGTGCTAATAGATCTTCTGATATGCAAGGAGATTT 1795
QY 758 CGGAGATTTTTTGGTAGATAGTAGAGTGCCTTCTGTTTTCACCTTCTTTACTCAG- 816

Db 1796 TTTGTTTTTTTGGTAGCTCAGTAGAAGTGGTCTCTGTTTTCGGCTTACTTTACTCAGA 1855
QY 817 CTGACTAGTGTCTCCCTTCGTTTCTAGTAAGTGGTGTAGAAATCACTGTCGGGCTT 876
Db 1856 GCTAGTAGTGTCTCCCTTAGTTTCTAGCAACTAGTGTGCAATCATGTGTGAGCTT 1915
QY 877 TACAGTTTTTAAACATATTATTCTGAAACATCACTGTTCTGCCAGAGTACCAACA 936
Db 1916 TAGGTTTTTAAACATATTATTAGATATTCTTAAACTATGAACCTTCTAACAGTACCAACA 1975
QY 937 CTGTCTATGTGATGATGCGCCCTCTAGACCTCACCACCGGACACATCTCCGCT 996
Db 1976 CTGTCTATGTGATGATGCTG---CCTCTAGACCTCGCCACGCTGACAGACGCTTCTCT 2032
RESULT 3
AK047969 2270 bp mRNA linear HFC 05-DEC-2002
LOCUS
DEFINITION
Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
library, clone:CI30025F01 product:RAS-related C3 botulinum
substrate 1, full insert sequence.
ACCESSION
VERSION AK047969.1 GI:26339063
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,

Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2270)

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saiki,D., Saichou,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,H., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

1. .2270
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C130025F01"
/db_xref="taxon:10090"
/clone="C130025F01"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
200. .778
/note="unnamed protein product; RAS-related C3 botulinum
substrate 1 (MG) [MGI:97845, GB|AK011072, evidence: BLASTN,
100%, match=987]
putative"
/codon_start=1
/protein_id="BAC33203.1"
/db_xref="GI:26339064"
/translation="MQAICVGVGVGAVGKTCLLISYTNAPFGEYIPVFNYSNV
WVGRPVNLGMDTQEDYDLPLSPYQDVELICFSLVSPAFENVRKAYEVR
HHCNPTPIILVCTKLDLDDKDTIEKLEKKLPIIYPOGLAWAKEIGAVKYLEUSAL
TORGLTVDEAIRAVLCPPPVKKRRKKCLLL"
2254. .2259
/note="putative"
2270
/note="putative"

polya_signal
polya_site
BASE COUNT 554 a 569 c 503 g 644 t
ORIGIN

Query Match 66.6%; Score 677.6; DB 11; Length 2270;
Best Local Similarity 85.3%; Pred. No. 2.9e-183;

	Matches	870;	Conservative	0;	Mismatches	119;	Indels	31;	Gaps	9;
QY	1	CCCCATTTCTTCTGCTCAGATTAAAGAAATTC	CCCAAAATACCTTGTGA	ACTAAGTTGCGTTG	60					
Db	1022	CCCCATTTCTTCTGCTCAGATTAAAGAAATTC	CCCAAAATACCTTGTGA	ACTAAGTTGCGTTG	1081					
QY	61	CTGAGAACCTAAGCTAAGCTGTTGAGAGA---	CTTTCCTCTAAGAA	AGCTGCGC	117					
Db	1082	CTGAGAACCTAAGCTAAGCTAAGCTGTTGAGAGA---	CTTTCCTCTAAGAA	AGCTGCGC	1141					
QY	118	TTCTGGGCTCAGGGGTGCAGACCTCCGTA	CTCCAGACCGTGTGA	CACAGACACGCC	177					
Db	1142	TTCTGGGCTCAGGGGTGCAGACCTCCGTA	CTCCAGACCGTGTGA	CACAGACACGCC	1201					
QY	178	TCCTTAATGACACCTGCCATGTAAACGAC	CTCTTAATGACACCTGCCATGTAAACGAC	CTCTTAATGAC	237					
Db	1202	TCCTTAATGACACCTGCCATGTAAACGAC	CTCTTAATGACACCTGCCATGTAAACGAC	CTCTTAATGAC	1261					
QY	238	AACCTTCTACTGTACGTACACCTGATGAG	GTGACAGCTCAGCTCTT	GATTTCATAGT	294					
Db	1262	AACCTTCTACTGTACGTACACCTGATGAG	GTGACAGCTCAGCTCTT	GATTTCATAGT	1321					
QY	295	AGTTCTCTAAATACACGCTGACCGGCTT	CTGCAGGCTTTGA	CAGAACTCTGCTCTG	354					
Db	1322	AGTTCTCTAAATACACGCTGACCGGCTT	CTGCAGGCTTTGA	CAGAACTCTGCTCTG	1381					
QY	355	TGTTGCTCTAAACGAAATTTCTTCTTAGT	CTGTGGTGTGCTGGTGGAGTGTGTGA	414						
Db	1382	TGTTGCTCTAAACGAAATTTCTTCTTAGT	CTGTGGTGTGCTGGTGGAGTGTGTGA	1439						
QY	415	ACAGAGCTCATCAAGGAGACAGACAGAT	TTTGA	CTAATATGAAGTATGATTAATTT	474					
Db	1440	ACAGAGCTCATCAAGGAGACAGACAGAT	TTTGA	CTAATATGAAGTATGATTAATTT	1499					
QY	475	ACACTCATTTGACAGGATTAATTC	AACTGAATAAAGTGTACGGGTAAAGCTTTTA	534						
Db	1500	ACACTCATTTGACAGGATTAATTC	AACTGAATAAAGTGTACGGGTAAAGCTTTTA	1559						
QY	535	ACGTTAAATTTCTGTCAAAACAGTAGAT	GACAAATGGCGATCTTATCAGTGTCTC	589						
Db	1560	ACGTTAAATTTCTGTCAAAACAGTAGAT	GACAAATGGCGATCTTATCAGTGTCTC	1619						
QY	590	-----TC	TGAGCCCCCTTCCCCTGCTGCTCCCTCCCAGATGGGGTGTGAGTCCA	642						
Db	1620	-----TC	TGAGCCCCCTTCCCCTGCTGCTCCCTCCCAGATGGGGTGTGAGTCCA	1679						
QY	643	TATTTAACTGGCCATCTC	CACAGTTGCTAACTAGCAAGTGTCTTTTAGACACCC	702						
Db	1680	TATTTAACTGGCCATCTC	CACAGTTGCTAACTAGCAAGTGTCTTTTAGACACCC	1737						
QY	703	TTCTTAACGAGCAATATGTGACCTGTACTA	TAAAGATCTTCTGATAATGCA-----TT	757						
Db	1738	TTCTTAACGAGCAATATGTGACCTGTACTA	TAAAGATCTTCTGATAATGCA-----TT	1797						
QY	758	CGGAGATTTTTTTGGTAGATAGTAGAAGT	CGGTTCTGTTCACCTTCTTTACTCAG-	816						
Db	1798	CGGAGATTTTTTTGGTAGATAGTAGAAGT	CGGTTCTGTTCACCTTCTTTACTCAG-	1857						
QY	817	CTGACTAGTGTCTCCCTTCGTTTCTAGTAA	CTGGGTGTAGAAATCACTGCTGCGGCTT	876						
Db	1858	GCTAGTGTGCTTCCCTTCGTTTCTAGTAA	CTGGGTGTAGAAATCACTGCTGCGGCTT	1917						
QY	877	TACAGTTTTTAAACATTTT	TAGATTTCTGAAACATCACTGTCTTGCCAGAGTACCACA	936						
Db	1918	TACAGTTTTTAAACATTTT	TAGATTTCTGAAACATCTTAAACATGTAGACCTTCTAAACAGTACCACA	1977						
QY	937	CTGTCAATGTGATGTGCGGCCCTCTAGAC	CTCACCACCGGACACATGCTTCCGCT	996						
Db	1978	CTGTCAATGTGATGTGCGGCCCTCTAGAC	CTCACCACCGGACACATGCTTCCGCT	2034						

RESULT 4
AK076023

LOCUS	AK076023	2308 bp	mRNA	linear	HTC 07-DEC-2002
DEFINITION	Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610100016 product:RAS-related C3 botulinum substrate 1, full insert sequence.				
ACCESSION	AK076023				
VERSION	AK076023.1 GI:26344957				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Iwata, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsumoto, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409 (6821), 685-690 (2001)				
MEDLINE	21085660				
PUBMED	11217851				
REFERENCE	5				
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
REFERENCE	6 (bases 1 to 2308)				
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,				
LOCUS	AK076023	2308 bp	mRNA	linear	HTC 07-DEC-2002
DEFINITION	Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610100016 product:RAS-related C3 botulinum substrate 1, full insert sequence.				
ACCESSION	AK076023				
VERSION	AK076023.1 GI:26344957				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Iwata, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsumoto, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409 (6821), 685-690 (2001)				
MEDLINE	21085660				
PUBMED	11217851				
REFERENCE	5				
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
REFERENCE	6 (bases 1 to 2308)				
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,				

Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numata, R., Onno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.					
Direct Submission					
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]					
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.					
Please visit our web site for further details.					
URL: http://genome.gsc.riken.go.jp/					
URL: http://fantom.gsc.riken.go.jp/					
Location/Qualifiers					
1..2308					
/organism="Mus musculus"					
/mol_type="mRNA"					
/strain="C57BL/6J"					
/db_xref="FANTOM_DB:2610100016"					
/db_xref="MGI:1898616"					
/db_xref="taxon:10090"					
/clone="2610100016"					
/tissue type="whole body"					
/clone_lib="RIKEN full-length enriched mouse cDNA library"					
/dev_stage="10 days embryo"					
239..817					
/note="unnamed protein product; RAS-related C3 botulinum substrate 1 (MGD MGI:97845, GI AK011072, evidence: BLASTN, 100%, match=987)"					
putative"					
/codon_start=1					
/protein_id="BAC36128.1"					
/db_xref="GI:26344958"					
/db_xref="MGI:97845"					
/translation="MQAIKCVVGVGAVGKTCILISYTNAPFGEYIPTFDNYSANVWVDCXPNVLGWDTAGOEDYDLRLPLSPQTDVFLICSLVSAASPNVRKWKYPEVRHCPNPIILVGTGLDLRDDKDTIEKLEKLTPTYPQGLAWAKEIGAVKYLECSALTGRLKTVFDEAIRAVLCPPPVKKRRKRCLLL"					
2293..2298					
/note="putative"					
2308					
/note="putative"					
BASE COUNT 559 a 580 c 518 g 650 t 1 others					
ORIGIN					
Query Match 64.9%; Score 660.2; DB 11; Length 2308;					
Best Local Similarity 84.8%; Pred. No. 3e-178;					
Matches 866; Conservative 0; Mismatches 123; Indels 32; Gaps 10;					
QY 1 CCCTATTCTGCTCAGATTAGAAATGCCAAATACCTTGCACTAGTTCGGTGTG 60					
Db 1061 CCCCCATTCTGCTCAGATTAAAAATGGCAAATACCTTCTGAATAGTTCGGTGTG 1120					
QY 61 CTGAGACACGTAAGCAGTAAGCTGTGTGAGAGA--CTTTGTCTTAAAGAGACTGCAGC 117					
Db 1121 CTGAGAACCTTAAGCAGCACTAACTCTCTTGAGAGACTTCTGTTGCTAAGAGACCGCAGC 1180					
QY 118 TTCTGG-GGTCAGGGGTGCAGACCCCTCCCGTAGCTCCAGACCGGTGTGCACGACGAGC 176					
Db 1181 TTCTGGAGCCAGGGGTGCAGACCCCTCCAGAGTTCACAGCCGTGTGCACAGCAGCAGC 1240					
QY 177 CTCCTTAATGACAGCGTGCCATGTAAAGCACCTGTACTTATCAGCCCATGCTCATACG 236					

```
Db 1241 CTCCTTATGACACGTTGCCATGTAAACGACCTGTAACTTATCAGCCCATGTTCTATTAG 1300
QY 237 TAACCTTTGACTGTACGTAC---GATGGGTGAACAGCTCTGTCTTTTGATTTTCATAGT 293
Db 1301 TAACCTTTGACTGTACGTACAGTGTACAGTGTACAGTCTGAGTCTTTGATTTTCATAGT 1360
QY 294 GAGTTCTCTAAATACACAGCTGACCGCTTCTGAGGCTTTGAACAGAACTCTGCTCTCT 353
Db 1361 GAGTTCTCTAAAGACACCGAGCTAGCTTTTGCAGACTTTTGAACAGAACTCTGCTCTCT 1420
QY 354 GTGTTGCTCTAAAGAAATTTCTTCTAGTCTGCTGCTGTGCTGCTGCTGAGTGTGA 413
Db 1421 GTGTTGCTCTAAAGAAATTTCTTCTAGTCTGCTGCTGTGCTGCTGAGTGTGA--GTGGA 1478
QY 414 AACAGAGCTCATCAAGAGGACAGACAGTATTTTGAATAATGAAGTAGAGTTAAT 473
Db 1479 AACAGAGCTCATCAAGAGGACAGACAGTATTTTGAATAATGAAGTAGAGTTAAT 1538
QY 474 TACACTACATTTGATGAGTATTTCACTGAATGAATGAAGTGTACCGGTAAAGCTTTT 533
Db 1539 TACACTACATTTGATGAGTATTTCACTGAATGAATGAAGTGTACCGGTAAAGCTTTT 1598
QY 534 AACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC--- 589
Db 1599 AACGGTTAATTTCTGTCAACAGTAGTGAATAATGGCCGATCTTATCAGTGTCTTGA 1658
QY 590 -----TCTGTAGCCCTTCTCCCGCTGTGCTGCTTCCCGATGGGGCTTGAAGTCC 641
Db 1659 GCATCCCTTTCGCGCCCTCCACCTCCACTGCGCTCCCTCCCGATGGGGCTTGAAGTCC 1718
QY 642 ATATTTAACTGGCCATCTCACAGTGTCTAACTAGCAAGTGTCTTTCTTAGACCCC 701
Db 1719 ATATTTAACTGGCCATCTCACAGTGTCTAACTAGCAAGTGTCTTTCTTTA-GAAGCC 1776
QY 702 CTTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAATGCA-----T 756
Db 1777 CTTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAAGCAGAGATT 1836
QY 757 TCGAGAGATTTTGGTAGATAGTAGAAGTGGTTCCTGTTTTCACTTCTTTACTCAG 816
Db 1837 TTTTGTGTTTTTGGTAGTCTAGTAGAAGTGGTTCCTGTTTTCACTTCTTTACTCAG 1896
QY 817 -CTGACTAGTCTTCCCTGCTTCTTCTAGTAACTGGGTGTAGAAATCAGTGTCTGGCT 875
Db 1897 AGTAGTAGTCTTCTTCTAGTAACTGGGTGTAGAAATCAGTGTCTGGCT 1956
QY 876 TTACAGTTTTTAACTATTTTAGATATTTGAAACATCACTGTCTGCGAGAGTACCAAC 935
Db 1957 TTACGGTTTTTAACTATTTTAGATATTTTAACTATGAACCTTTTAAACAGTACCAAC 2016
QY 936 ACTGTCTAGTGTATGATGCGCCCTCTAGACCTCAACCGCGGACAGATGCTTCGG 995
Db 2017 ACTGTCTAGTGTATGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2073
QY 996 T 996
Db 2074 T 2074

RESULT 5
BU514423
LOCUS
DEFINITION BU514423 978 bp mRNA linear EST 12-SEP-2002
AGENCOURT 10119699 NIH MGC 134 Mus musculus cDNA clone
IMAGE:6511045 5', mRNA sequence.
ACCESSION BU514423
VERSION BU514423.1 GI:22821949
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 978)
```

AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILLNL at:
<http://image.llnl.gov>
Plate: LUM14080 row: o column: 14
High quality sequence start: 17
High quality sequence stop: 714.
Location/Qualifiers

FEATURES
source

1. 978
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6511045"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT 249 a 224 c 201 g 301 t 3 others
ORIGIN

Query Match 58.4%; Score 594.2; DB 13; Length 978;
Best Local Similarity 85.8%; Pred. No. 1.9e-159;
Matches 762; Conservative 0; Mismatches 94; Indels 32; Gaps 8;

```
QY 1 CCCCTATCTGCTCAGATTAAAGATTGCCAAATACCTTGTGAACTAAGTGGTGTG 60
Db 70 CCCCTATCTGCTCAGATTAAAGATTGCCAAATACCTTGTGAACTAAGTGGTGTG 129
QY 61 CTGAGAACAGCTAAGCACTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGAGCTGCAGC 117
Db 130 CTGAGAACACTAAGCACTAAGCTGTGTGAGAGACTTCTGTCTAAGAGACCGAGC 189
QY 118 TCTGGGCTCAGGGGTGACAGACCTCCCTGCTAGCTCCAGAGCCGCTGTGACACAGACAGCC 177
Db 190 TCTGGAGCAGGGGTGACAGACCTCCCTGCTAGCTCCAGAGCCGCTGTGACACAGACAGCC 249
QY 178 TCCTTAATGACAGCTGCCATGTAAACCACTGTAACTTATCAGCCCATGCTCATAGT 237
Db 250 TCCTTAATGACAGCTGCCATGTAAACCACTGTAACTTATCAGCCCATGCTCATAGT 309
QY 238 AACTTTGTACTGTACGTCAC---GATGGGTGAACAGCTCTGCTCTTTGATTTTCATAGT 294
Db 310 AACTTTGTACTGTACGTCACAGTGTGAGTGTGACAGCTCAGCTCTTTGATTTTCATAGT 369
QY 295 AGTTCTCTAAATACAGCTGACCGCTTCTGAGGCTTTTGAACAGAACTCTGGCTCCTG 354
Db 370 AGTTCTCTAAAGACACCGAGCTAGCTTTTGCAGACTTTTGAACAGAACTCTGGTTCCTG 429
QY 355 TGTGCTCTCTAAGAAAGTATCTGTTCTCTAGTCTGCTGGGTGTGCTGGGTGAGTGTGAA 414
Db 430 TGTGCTCTCTAAGAAAGTATCTGTTCTCTAGTCTGCTGGGTGTGCTGGGTGAGTGTGAA 487
QY 415 ACAGCAGCTCATCAAGAGGACAGACAGTATTTTGAATAATGAAGTAGAGTAAAGCTTTTA 474
Db 488 ACAGCAGATGATCAAGAGGACAGACAGTATTTTGAATAATGAAGTAGAGTAAAGCTTTT 547
QY 475 ACACACTATGTACATGGAGTAAATTCACACTGAATAAAGTGTCAACGGTAAAGCTTTTA 534
Db 548 ACACACTATGTACATGGAGTAAATTCACACTGAATAAAGTGTCAACGGTAAAGCTTTTA 607
QY 535 ACGTTAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC----- 589
```

Db 608 ACGGTTAAATTTCTGTCAAAACAGTAGGTGACAAATGGCCGATCTTATCAGTGTCTCTTGAG 667
 QY 590 -----TCTTGAGGCCCCCTTCCCTGCTGCTCCCGAGATGGGCGGTGAGTCCA 642
 Db 668 CATCCCTTCGCGCCCCCACCCTCCACTCCGCTCCCTCCCGAGATGGGCGGTGAGTCCA 727
 QY 643 TATTTAAACTGGCCATCTCCACAGTTCCTAACTAGCAAGTGTCTTTCTTTAGGACCCCC 702
 Db 728 TATTTAAACTGGCCATCTCCACAGTTCCTAACTAGCAAGTGTCTTTCTTTAG--ACCC 785
 QY 703 TTTCTAACGAGCAATATGCTGACCTGTACTAAGATCTTTCTGATAAAGCA-----T 756
 Db 786 TTTCTTACGAGCAATATGCTGACCTGTACTAAGATCTTTCTGATAAAGCAAGATTT 845
 QY 757 TCGGAGATTTTTTGTGTAGAT--AGTAGAGTGGCTTCTGTTTCAACCTTCTTACTC 814
 Db 846 TTTGTTTTTTTTTGTGTAGTCAAGTAGAGTGGCTTCTGTTTCTGCGCTACTTTTACTC 905
 QY 815 --AGCTGACTAGTGTCTCCCTTCTGTTTCTAGTAACTGGGTGTAGAAA 860
 Db 906 AGAGCTAGTGTAGTGGTCTCTTCTAGTAACTGGGTGTAGAAA 953

RESULT 6
 CA318577
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
 NIH-MGC http://img.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaaps.r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library prepared by: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.
 Location/Qualifiers
 1..826
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6813639"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="PH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP FWO"
 /note="Organ: Brain; Vector: pyx-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCGAGACAG. This library was created for the University

FEATURES

source

BASE COUNT 226 a 195 c 162 g 242 t 1 others
 ORIGIN
 Query Match 55.3%; Score 562; DB 14; Length 826;
 Best Local Similarity 89.7%; Pred. No. 3 le-150;
 Matches 677; Conservative 0; Mismatches 56; Indels 22; Gaps 6;
 QY 1 CCCCTATTCTTCTCAGATTAAAGAAATGCCAAATACCTTGTCAATAAGTTGCGTTGTG 60
 Db 76 CCCCCTATTCTTCTCAGATTAAAGAAATGCCAAATACCTTGTCAATAAGTTGCGTTGTG 135
 QY 61 CTGAGAACAGTAAGCACTAAAGCTGTTGAGAGA---CTTTGTCCCTTAAAGAACTCCAGC 117
 Db 136 CTGAGAACAGTAAGCACTAAAGCTGTTGAGAGACTTCTTTGAGAGACTTCTTTGAGAGACCCGAGC 195
 QY 118 TTCTGGGCTCAGGGGTGCAGACCCCTCCCTAGTCTCCAGACCGTGTGACACAGCAGCC 177
 Db 196 TTCTGGAGCCAGGGGTGCAGACCCCTCCAGAGTTCCAGACCGTGTGACACAGCAGCC 255
 QY 178 TCTTAAATGACAGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATACGT 237
 Db 256 TCTTAAATGACAGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATACGT 315
 QY 238 AACTTTGTAAGTACGTAC--GATGGGTGTAAAGCTCTGCTCTTCAATTTCAATGATG 294
 Db 316 AACTTTGTAAGTACGTACGTACGTGAGTGTGACAGCTCAGCTCTTTGATTTCAATGATG 375
 QY 295 AGTTCTCTAAATATACAGCTGACCGGCTTCTGAGAGCTTTGAACAGAACTCTGGCTCCTG 354
 Db 376 AGTTCTCTAAAGACAGCGGCTAGCTTTTTCAGAGCTTTGAACAGAACTCTGGTCCG 435
 QY 355 TGTGCTCTTAACGAAGTATTCGTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTGAA 414
 Db 436 TGTGCTCTTAAGTATTCGTTCTAGTGTGGGTGTGCTGGGTGGA--GTGTGAA 493
 QY 415 ACAGAGCTCATCAAGGAGAGACAGAGTATTTTGAATATATGAAGTATGAGATTAATTT 474
 Db 494 ACAGAGCTCATCAAGGAGAGACAGAGTATTTTGAATATATGAAGTATGAGATTAATTT 553
 QY 475 ACACATACATTTACATGAGTAACTCACTGAATAAAGTGTACGGGTAAAGCTTTTAA 534
 Db 554 ACACATACATTTACATGAGTAACTCACTGAATAAAGTGTACGGGTAAAGCTTTTAA 613
 QY 535 AGGTTAAATTTCTGTCAAACAGTAGATGACAAATGCCCGCATCTTATCAGTGTCTC----- 589
 Db 614 ACGTTTAAATTTCTGTCAAACAGTAGATGACAAATGCCCGCATCTTATCAGTGTCTCTGAG 673
 QY 590 -----TCTTGAGCCCCCTTCCCGCTGCTGCCCTCCCGAGATGGGCGTTGAGTCCA 642
 Db 674 CATCTCTTGGCGCCCCCACCCTCCACTGCGCTCCCTCCCGAGATGGGCGTTGAGTCCA 733
 QY 643 TATTTAACTGGCCATCTCCTCACAGTTGCTAACTTAGCAAGTGTCTTTCTTTAGACCCCC 702
 Db 734 TATTTAACTGGCCATCTCCTCACAGTTGCTAACTTAGCAAGTGTCTTTCTTTTAA--GAACCCC 791
 QY 703 TTTCTAACGAGCAATATGCTGACCTGTACTATAA 737
 Db 792 TTTCTAACGAGCAATATGCTGAGCTGTACTATAA 826

RESULT 7

BQ177713

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

BQ177713 751 bp mRNA linear EST 30-APR-2002
 UI-M-ERO-bwp-j-03-0-UI.r1 NIH_BMAP_ERO Mus musculus cDNA clone
 IMAGE: 5700194 5', mRNA sequence.
 BQ177713
 BQ177713.1 GI:20353205
 EST.
 Mus musculus (house mouse)

```

ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 751)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. 751
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 5700194"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (TI phage resistant)"
/clone_lib="NIH BMAP ERO"
/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GTCGCTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). "Gene Discovery in the Developing Mouse Nervous
System", supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 198 a 180 c 154 g 217 t 2 others
ORIGIN
Query Match 54.9%; Score 558.2; DB 13; Length 751;
Best Local Similarity 89.6%; Pred. No. 3.7e-149;
Matches 673; Conservative 0; Mismatches 55; Indels 23; Gaps 6;
QY 1 CCCATTCTTCTGCTAGATTAGAAATGCGAAATACCTTGTGAACCTAAGTTCGGTTGTG 60
Db 6 CCCCCATTCTGCTAGATTAAAAATGCGAAATACCTTGTGAACCTAAGTTCGGTTGTG 65
QY 61 CTGAGAACACGTAAGCACTAAGCTGTTGAGAGA---CTTGTCTTAAAGAACTGCAGC 117
Db 66 CTGAGAACACCTAAGCACTAAACTCTCTTGAAGAGACTTCTGTGCTAAGAACCGCAGC 125
QY 118 TTCTGGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGCTGTGACACAGCAGCC 177
Db 126 TTCTGGAGCCAGGGGTGCAGACCCCTCCAGAGTTCACAGCCGTGTGACACAGCAGCC 185
QY 178 TCTTAAATGACACGCTGCCATGTAAGCGACCTGTAACTTATCAGCCCATCTCAATACGT 237
Db 186 TCTTAAATGACACGCTGCCATGTAAGCGACCTGTAACTTATCAGCCCATCTCAATACGT 245
QY 238 AACTTTGTACTGTACGTAC---GATGGGTGTAAGAGCTCTGCTCTTGAATTCATGTG 294
Db 246 AACTTTGTACTGTACGTACAGTGATGAGTGTGACAGCTCAGCTCTTGAATTCATGTG 305
QY 295 AGTTCTCTAAAATACCAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTGGCTCCTG 354

```

```

Db 306 AGTTTCTAAAAGACGACCGCACTAGCTTTTCAGACTTTGAACAGAACTCTGGTTCCTG 365
QY 355 TGTTCCTCTAACGAAGATTCTGTTCCTAGTCGTTGGGTGCTGCTGGTGGAGTGTGTGAA 414
Db 366 TGTTCCTCTAACGAAGATTCTGTTCCTAGTGTGGGTGCTGCTGGTGGAGTGTGTGAA 423
QY 415 ACACGACGCTCATCAAGGAGACAGACAGTATTTTCTACTAATATGAAGTAGAGTAATTT 474
Db 424 ACACGACATGATCAAGGAGACAGACAGTATTTTCTACTAATATGAAGTAGAGTACTTT 483
QY 475 ACACCTAGATTGACATGAGGTAATTTCACTGAATAAAGTGTACGGGTAAAGCTTTTAA 534
Db 484 ACACCTAGATTGACATGAGGTAATTTCACTGAATAAAGTGTACGGGTAAAGCTTTTAA 543
QY 535 ACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC----- 589
Db 544 ACGGTTAATTTCTGTCAACAGTAGTGTGACAAATGGCCGATCTTATCAGTGTCTCTTGAG 603
QY 590 -----TCTTGAAGCCCTTCCCTGCTGTGCTCCCTCCAGATGGGGCGTTGAGTCCA 642
Db 604 CATCCCTTGGCGCCCGCCACCTCCACTCCGCTCCTCCAGATGGGGCGTTGAGTCCA 663
QY 643 TATTAACTGGCCCATCTCTACAGTGTCTAACTTAGCAAGTGTCTTTTAGGACCCCC 702
Db 664 TATTAACTGGCCCATCTCTACAGTGTCTAACTTAGCAAGTGTCTTTTAGGACCCCC 720
QY 703 TTCTTAACGAGCAATATGTCTGACCTGTACT 733
Db 721 TTCTTAACGAGCAATATGTCTGACCTGTACT 751
RESULT 8
BUSE11766 960 bp mRNA linear EST 12-SEP-2002
AGENCOURT 10114251 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6506583 5', mRNA sequence.
BUSE11766
BUSE11766.1 GI:22817999
Mus musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 960)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14069 row: e column: 16
High quality sequence stop: 669.
Location/Qualifiers
1. 960
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6506583"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 134"
/note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI. Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT 246 a 218 c 196 g 298 t 2 others

```


ORIGIN	Query Match	54.3%;	Score	552.2;	DB	13;	Length	960;
	Best Local Similarity	86.6%;	Pred. No.	2.3e-147;				
	Matches	733;	Conservative	0;	Mismatches	79;	Indels	34;
	Gaps	10;						
QY	45	ACTAAGTGGCTGTGGCTGAGACACGTAAGCACTAAGCTGTGAGAGAC---	TTTGCC	101				
Db	3	AATAAGTGGCTGTGGCTGAGAACACCTAAGCACTAAATCTCTTTGAGAGACTTCTGTG	62					
QY	102	TTAAGACACTGACGCTTCTGGGCTCAGGGTGACAGCCCTCCCGTAGCTCCACAGCGT	161					
Db	63	CTAAGACACCGACGCTTCTGAGCCAGGGTGACAGCCCTCCAGAGTTCCACAGCGT	122					
QY	162	GTGACACAGCAGCGCTCTTAATACACAGCGTCCCATGTAAGCAGCTGTAACTTATCAG	221					
Db	123	GTGACACAGCAGCGCTCTTAATGACAGCTGTCCATGTAAAGCAGCTGTAACTTATCAG	182					
QY	222	CCCATGCTCATTAACCTTGTACTGTACGTACCTAC---GATGGGTGTAACAGCTCTGCT	278					
Db	183	CCCATGCTCATTAACCTTGTACTGTACGTACGTACGTACGTACGTACGTACGTACGT	242					
QY	279	CTTTGATTTCAAGTGTCTCTAAATACCAGCTGACCGCTTCTGACGCTTTGAAC	338					
Db	243	CTTTGATTTCAAGTGTCTCTAAAGACAGCGGCTAGCTTTTGACAGCTTTGAAC	302					
QY	339	AGAACTCTGGCTCTGCTGCTCTTAACGAAAGTATCTGCTTCTGCTAGCTGCTGCTGCT	398					
Db	303	AGAACTCTGGCTCTGCTGCTCTTAATGAAGTATCTGCTTCTGCTAGCTGCTGCTGCT	362					
QY	399	GGGTGGAGTGTGAACACGAGCTCATCAAGAGGACAGACAGATTTTGTACTAATATG	458					
Db	363	GGGTGGA--GTGTGAACACGAGCTCATCAAGAGGACAGACAGATTTTGTACTAATATG	420					
QY	459	AAGTAGAGATTAATTTACACTACATTTGATGAGTAAATCAACTGAAATAAAGTGTCA	518					
Db	421	AAGTAGAGATTAATTTACACTACATTTGATGAGTAAATCAACTGAAATAAAGTGTCA	480					
QY	519	CGGTAAGCTTTTAAACGGTTAAATTTCTGTCAACAGTAGATGACAAATGGCGATCTT	578					
Db	481	CGGTAAGCTTTTAAACGGTTAAATTTCTGTCAACAGTAGATGACAAATGGCGATCTT	540					
QY	579	ATCAGTGTCTC-----TCTTGAGCCCTCTTCCCTCTGCTGCTTCCCTCCACAG	626					
Db	541	ATCAGTGTCTCTTGAGATCTCTTCCGCGCCCTCCCTCCCTCCCTCCCTCCCTCC	600					
QY	627	TGGGGGCTTGAGTCCATATTTAACTGGCCATCTCCACAGTTGCTTAACTAGCAAGTGT	686					
Db	601	TGGGGGCTTGAGTCCATATTTAACTGGCCATCTCCACAGTTGCTTAACTAGCAAGTGT	659					
QY	687	TTTCTTTAGACCCCTCTTACAGGCAATATGCTGACCTGTACTATTAAGATCTTCT	746					
Db	660	TTTCTTTA-GAACCCTCTTAAAGCAATATGCTGACCTGTACTATTAAGATCTTCT	718					
QY	747	GATAAGTCA-----TTCGGAGATTTTGTGTAGATAGTAGAAGTGCCTTCTG-TTT	798					
Db	719	GATAAGTCA-----TTCGGAGATTTTGTGTAGATAGTAGAAGTGCCTTCTG-TTT	778					
QY	799	TCACCTTCTTTACTC---AGCTGACTAGTGTCTTCCCTCGTTT-TCCTAGTACGGGTG	854					
Db	779	TCGCTTTACTTTACTCAGAGCTAGTTAGTGTCTTCCCTAGTCTTCCCTAGCACTAGGTG	838					
QY	855	TAGAAA	860					
Db	839	TGCAAA	844					
RESULT 9	BG65146	557 bp	mRNA	linear	EST	30-APR-2001		
LOCUS	BG65146							
DEFINITION	DRABZ12 Rat DRG Library Rattus norvegicus cDNA clone DRABZ12 5',							
ACCESSION	BG65146							

VERSION	BG65146.1	GI:13887068
KEYWORDS	EST.	
SOURCE	Rattus norvegicus (Norway rat)	
ORGANISM	Rattus norvegicus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
REFERENCE	1. (bases 1 to 557)	
AUTHORS	Xiao, H.S., Huang, Q.H., Zhang, F.X., Bao, L., Lu, Y.J., Guo, C., Yang, L.,	
	Huang, W.J., Fu, G., Xu, S.H., Cheng, X.P., Yan, Q., Zhu, Z.D., Zhang	
TITLE	X., Chen, Z., Han, Z.G. and Zhang, X.	
JOURNAL	Identification of gene expression profile of dorsal root ganglion	
MEDLINE	in the rat peripheral axotomy model of neuropathic pain	
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)	
COMMENT	22056133	
	12060780	
	Contact: Zhang Xu	
	Laboratory of Sensory System	
	Institute of Neuroscience	
	320 Yue Yang Road, Shanghai 200031, P.R.China	
	Tel: 86-21-64748700-121	
	Fax: 86-21-64713446	
	Email: xu.zhang@ion.ac.cn	
	This clone is also available at Chinese National Human Genome	
	Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,	
	Pudong New Area, P.R.China. Please contact with Zhang Xu	
	(xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)	
	PCR Primers	
	FORWARD: T3	
	BACKWARD: T7	
	Seq primer: T3	
	POLYA=No.	
FEATURES	Location/Qualifiers	
	1..557	
	/organism="Rattus norvegicus"	
	/mol_type="mRNA"	
	/strain="Sprague-Dawley"	
	/db_xref="taxon:10116"	
	/clone="DRABZ12"	
	/sex="male"	
	/tissue_type="dorsal root ganglion"	
	/dev_stage="adult"	
	/clone_lib="Rat DRG Library"	
	142 a 130 c 128 g 156 t 1 others	
BASE COUNT		
ORIGIN		
Query Match	54.2%;	Score 551.2; DB 10; Length 557;
Best Local Similarity	99.3%;	Pred. No. 3.2e-147;
Matches	553;	Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	55	GTGTGCTGAGAACAGCTAAGCACTAAGCTGTGTGAGAGACTTTGTCTTAAGAGACTGC 114
Db	1	GTGTGCTGAGAACAGCTAAGCACTAAGCTGTGTGAGAGACTTTGTCTTAAGAGACTGC 60
QY	115	AGTTCTGGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGGTGCACAGCACA 174
Db	61	AGTTCTGGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGGTGCACAGCACA 120
QY	175	GCCTCCTTAATGACAGCTGCCATGTAAACGCACTGTAACTATCAACCCATGTCTATTA 234
Db	121	GCCTCCTTAATGACAGCTGCCATGTAAACGCACTGTAACTATCAACCCATGTCTATTA 180
QY	235	CGTAACCTTTGCTACTGTACGTCACGATGGGTGTAAACAGCTCTGCTCTTTGATTTCAATG 294
Db	181	CGTAACCTTTGCTACTGTACGTCACGATGGGTGTAAACAGCTCTGCTCTTTGATTTCAATG 240
QY	295	AGTTCTCTAAATACACAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTTGCTCTG 354
Db	241	AGTTCTCTAAATACACAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTTGCTCTG 300
QY	355	TGTTGCTCTTAAACGAACTATTCTGCTTCTTCTAGTCTGCTGCTGCTGCTGCTGCTGCTG 414
Db	301	TGTTGCTCTTAAACGAACTATTCTGCTTCTTCTAGTCTGCTGCTGCTGCTGCTGCTGCTG 360

QY 415 ACACGAGTCATCAAGAGCAGACAGACAGTATTTGACTAATATGAAGTAGAGATTAAATTT 474
 Db 361 ACACGAGTCATCAAGAGCAGACAGACAGTATTTGACTAATATGAAGTAGAGATTAAATTT 420
 QY 475 ACACCTACATTTGACATGAGTAAATCAACTGAATAAAAAGTGTCAAGGGTAAAGCTTTTAA 534
 Db 421 ACACCTACATTTGACATGAGTAAATCAACTGAATAAAAAGTGTCAAGGGTAAAGCTTTTAA 480
 QY 535 ACAGTTAATTTCTGTCAAAACAGTAGATGACAAATGGCCGATCTTATCAAGTGTCTCTCTTG 594
 Db 481 ACAGGTAAATTTCTGTCAAAACAGTAGATGACAAATGGCCGATCTTATCAAGTGTCTCTCTTG 540
 QY 595 AGCCCCCCTTCCCTCG 611
 Db 541 AGCCCCCCTTCCCTCG 557

RESULT 10
 CA319273
 LOCUS
 DEFINITION UI-M-FW0-cbz-a-24-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
 IMAGE: 6816385 5', mRNA sequence.
 CA319273
 VERSION CA319273.1 GI:24537397
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 793)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.
 Location/Qualifiers
 1. 793
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6816385"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5, 14.5, 15.5, 16.5, 17.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_FW0"
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCAGACAG. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

200 a 185 c 165 g 242 t 1 others

BASE COUNT

ORIGIN
 Query Match 52.8%; Score 537.4; DB 14; Length 793;
 Best Local Similarity 87.2%; Pred. No. 3.7e-143;
 Matches 677; Conservative 0; Mismatches 72; Indels 27; Gaps 7;
 QY 64 AGAACACGTAAGCCTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGACAGTGCAGCTTC 120
 Db 1 AGAACACCTAAGCCTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGACAGTGCAGCTTC 60
 QY 121 TGGGCTCAGGGGTGAGACACCTCCGCTAGCTCCAGACCGGTGTGACACAGACACGCTCC 180
 Db 61 TGGAGCCAGGGGTGAGACACCTTCCAGAGTCCAGACCGGTGTGACACAGACACGCTCC 120
 QY 181 TTAATGACACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGTAAC 240
 Db 121 TTAATGACACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGTAAC 180
 QY 241 TTTGTACTGTAGCTCAC---GATGGGTGTAAACGCTCTGTCTTTGATTTTATAGTAGT 297
 Db 181 TTTGTACTGTAGCTCACAGTGTAGTGTGACGCTCAGCTCTTTGATTTTATAGTAGT 240
 QY 298 TCTCTAAATACACGCTGACCGGCTTCTGAGGCTTTTGAACAGAACTCTGGCTCCTGTGT 357
 Db 241 TTTCTAAAGACACCGGCTAGCTTTTGCAGACTTTTGAACAGAACTCTGGTCTCTGTGT 300
 QY 358 TGCCTCTAAACGAAGTATTTCTTCTTCTAGTCTGGGTGTGTGGGTGGAGTGTGAAACA 417
 Db 301 TGCCTCTAAACGAAGTATTTCTTCTTCTAGTCTGGGTGTGTGGGTGGAGTGTGAAACA 358
 QY 418 CGAGTCTAAAGAGACAGACAGTATTTTGCATATATGAAGTAGAGATTAATTACA 477
 Db 359 CGAGTCTAAAGAGACAGACAGTATTTTGCATATATGAAGTAGAGATTAATTACA 418
 QY 478 CTACATTTGATCATGAGTAATTTCAACTGAATAAAGTGTCAAGGTGAAGCTTTTAAAG 537
 Db 419 CTACATTTGATCATGAGTAATTTCAACTGAATAAAGTGTCAAGGTGAAGCTTTTAAAG 478
 QY 538 GTTAATTTCTTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC----- 589
 Db 479 GTTAATTTCTTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCCTTGAGCAT 538
 QY 590 ----TCTTGAGCCCTTCTCCCTCTGCTGCTCCCTCCAGATGGGGCTGTGAGTCCATAT 645
 Db 539 CCCTTCGGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCATAT 598
 QY 646 TTAACCTGGCCATCTCACAGTGTCTACTTAGCAAGTGTCTTTTCTTAGGACCCCTTC 705
 Db 599 TTAACCTGGCCATCTCACAGTGTCTACTTAGCAAGTGTCTTTTCTTAGGACCCCTTC 656
 QY 706 TTAACGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTGATAATGCA-----TTCGG 760
 Db 657 TTAACGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTGATAATGCA-----TTCGG 716
 QY 761 AGATTTTGTGTAGATAGTAGAGTGGCTTCTGTTTCCCTTCCCTTCTTACTACTAG 816
 Db 717 GTTTTGTGTGTAGTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 772

RESULT 11
 BI646446
 LOCUS 603279026f1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5319127 5',
 DEFINITION mRNA sequence.
 BI646446
 ACCESSION BI646446.1 GI:15560682
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 897)
 NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LLM11809 row: d column: 08 High quality sequence stop: 890.

FEATURES source

```

1. .897
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129, C57BL/6J, FVB/N"
/db_xref="taxon:10090"
/clone="IWAGR:5319127"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam3"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics

```

BASE COUNT	228 a	213 c	190 g	266 t
ORIGIN				

Query Match	52.8%;	Score 535.2;	DB 12;	Length 897;
Best Local Similarity	83.6%;			
Pred. No. 1.7e-142;				
Matches 734;	Conservative	0;	Mismatches 113;	Indels 31;
Gaps 10;				

60	QY	GTGAGAAACAGCTAAGCACTAAGCTGTGAGAGA---CTTTGTCTTAAAGAGACTGCAG	116
1	Db	GTGAGAAACACCTAAGCACTAAACTCTTTGAGAGACTTCTGTCTTAAAGAAGCCGACG	60
117	QY	CTTCTGGGCTCAGGGGTGCAGACCCCTCCGTAGCTCCAGACCGTGTGCACACGACAGC	176
61	Db	CTTCTGGAGCCA---GGGTGCAGGCCCTTCAGAGTCCAGACCGTGTGCACACGACAGC	119
177	QY	CTCCTTTAATGACACGCTGGCATGTAAACGCAACCTGTAACTTATCAGGCCATGCTCATPACG	236
120	Db	CTCCTTTAATGACACGTTGCCATGTAACGGCACCTGTAACTTATCAGGCCATGTTCAATPACG	179
237	QY	TAACTTTGTACTCTPACGTCAAC---GATGGGTGTAAACAGCTCTGCTCTTTTCATTTTCATAGT	293
180	Db	TAACTTTGTACTGTACGTACAGTGATGATGTGCACAGCTCAGCTCTTTGATTTTCATAGT	239
294	QY	GAGTTCTCTAAATACACAGCTGACCGGCTTCTCGAGGCTTTTGAAACGAACTCTGGCTCCT	353
240	Db	GAGTTTCTTAAAGACACGCGGACTAGCTTTTGCAGACTTTTGAACGAACTCTGGTTCCCT	299
354	QY	GTGTTGCCCTTAACGAAGTATTCGTGTTCTTAGTCGTGGGTGTGCTGGGTGGAGTGTGA	413
300	Db	GTGTTGCCCTTAATGAAGTATTCGTGTTCCAGTTGTGGGTGTGCTGGGTGGAGTGTG--A	357
414	QY	AACACGACGCTCATCAAGGAGACACAGCAGTATTTTGACTAATATGAAGTAGAGATTAATT	473
358	Db	AACACGACATGATCAAGGAGACACAGCAGTATTTTGACTAATATGAAGTAGAGATTAATT	417
474	QY	TACACTACATGTTACATGGAGTAATTCMACTGAATAAAGTGTCACGGTAAAGCTTTTT	533
418	Db	TACACTACATGTTACATGGAGTAATTCMACTGAATAAAGTGTCACGGGTAAAGCTTTTT	477
534	QY	AACGGTTAATTTCTGTCAACACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC---	589
478	Db	AACGGTTAATTTCTGTCAACACAGTAGTAGCAAAATGGCCGATCTTATCAGTGTCTTTGA	537

590	QY	-----TCTTGAGCGCCCGCTCCCGCTGCTGCTCCCTCCCGAGATGGGCGTGTAGTCC	841
598	Db	GCATCCCTTCGGGCGCCCGCAACCTCCACGTGCGTCCCTCCCGAGATGGGCGGTGAGTCC	597
642	QY	ATATTAAACT--GGCCATCCTCACAGTTCCTAACTTAGCAAGTGCCTTTCTTTAGGACCC	700
598	Db	ATATTAAACTGGGCCACCTCACAGTGTCTAACTTAGCAAGTGCCTTTCTTTA-GAAACC	656
701	QY	CCTTCTTAACGAGCAATGTCTGACCTGCTACTATAAGATCTTCTGATAATGCA---T	756
657	Db	CCTTCTTAACGAGCAATATGTCTGACCTGCTACTATAAGATCTTCTGATAACGAGA	716
757	QY	TCGGAGATTTTTTGGTAGATAGTAGAAGTGGCTTCTGTTTCACTCCCTTTACTCAG	816
717	Db	TTTGGTCCCTCTGGTAGCTCAGTAGAAGTGGCTGTTTTCGCTTACTTTACTCAG	776
817	QY	-CTGACTAGTGTCCCTCGTTTCTTAGTAAGTGGGTGT---AGAAATCAGTGTGCG	872
777	Db	AGCTAGTTAGTGTCCCTTAGTTTTCTAGCCAACTAGAGTGTGCCAAATCATGTGTGCG	836
873	QY	GCTTACAGTTTTTAAACTATTTTAGATATTCGAAAC	910
837	Db	CGCCTTACGGGTTTAACTATTTTAGATATTCCTTAAC	874

RESULT 12

BI851489
LOCUS
DEFINITION

714 bp . mRNA
EST 10-OCT-2001
603377695Fi NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:539207 5',
mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BI851489
BI851489
BI851489.1 GI:15992236
EST.
Mus musculus
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 714)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.

```

FEATURES
source
high quality sequence stop: /13.
Location/Qualifiers
1. .714
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5390207"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
187 a 172 c 149 g 206 t
BASE COUNT

```

Query Match	52.0%;	Score 528.6;	DB 12;	Length 714;
Best Local Similarity	89.5%;	Pred. No. 1.2e-140;		
Matches 642;	Conservative	0;	Mismatches 54;	Indels 21;
				Gaps 6;

QY 2 CCTATTCTTCTGCTCAGATTAGATTCGCAAAATACCTTGTGAACCTAAGTTGGTTGTC 61
Db 1 CCCCATTTCTGCTCAGATTAGATTCGCAAAATACCTTGTGAACCTAAGTTGGTTGTC 60
QY 62 TGAGAACACCTGATAGCCTAAGCTCTTGAGAGA---CTTTGTCTTAAAGAGACTGACGT 118
Db 61 TGAGAACACCTGATAGCCTAAGCTCTTGAGAGACTCTTGTTGTTAAGAGACCGCAGCT 120
QY 119 TCTGGGCTCAGGGGTGAGACCCCTCCGTAAGTCCGAGACCGGTGTGACAGACAGCCCT 178
Db 121 TCTGGAGCCA-GGGTGCAGACCCCTCCGTAAGTCCGAGACCGGTGTGACAGACAGCCCT 179
QY 179 CCTTAATGACACGCTGCATGTACGACCTGACCTGTAACCTATACAGCCCATGCTCAATAGTA 238
Db 180 CCTTAATGACACGCTGCATGTACGACCTGTAACCTATACAGCCCATGCTCAATAGTA 239
QY 239 ACTTTGACTGTAGCTCAC---GATGGGTGTAAAGCTCTGCTCTTTGATTTCATAGTA 295
Db 240 ACTTTGACTGTAGCTCACAGTGTAGTGTGACAGCTCAGCTCTTTGATTTCATAGTA 299
QY 296 GTTCTCTAAATACCACTGACCGGCTTTCGACGCTTTGAACAGAACTCTGGCTCCTGT 355
Db 300 GTTTTCTAAAGACACCGCAGCTAGCTTTTGCAGACTTTGAACAGAACTCTGGTCTCTGT 359
QY 356 GTTGGCTCTAACGAAGTATTCTGTCTCTAGTCTGCTGGGTGTGCTGGGTGGTGTGTA 415
Db 360 GTTGGCTCTAACGAAGTATTCTGTCTCTAGTCTGCTGGGTGTGCTGGGTGGTA--GTTGTA 417
QY 416 CAGACGCTCATCAAGGAGACAGACAGTATTTTGAACATAATGATAGAGATTAATTA 475
Db 418 CAGACATGATCAAGGAGACAGACAGTATTTTGAACATAATGATAGAGATTAATTA 477
QY 476 CACTACATTTGATGAGTAATTCACATGAATAAAGTGTACCGGTAAGCTTTTAA 535
Db 478 CACTACATTTGATGAGTAATTCACATGAATAAAGTGTACCGGTAAGCTTTTAA 537
QY 536 CGGTAAATTTCTGCAACAGTAGATACACAAATGGCGATCTTATCAGTGCT----- 588
Db 538 CGGTAAATTTCTGCAACAGTAGATACACAAATGGCGATCTTATCAGTGCTTTGAGC 597
QY 589 -----CTTCTGAGCCCTTCCCTCTGCTGCTCCCTCCGATGGGCGTTGAGTCCATA 644
Db 598 ATCCCTCGGGCCCCCACCCTCCTGCTGCTCCCTCCGATGGGCGTTGAGTCCATA 657
QY 645 TTTAACTGGCCACCTCACAGTT-CTAACTTAGCAAGTGCTTTCTTTAGAACCCC 701
Db 658 TTTAACTGGCCACCTCACAGTT-CTAACTTAGCAAGTGCTTTCTTTAGAACCCC 713

RESULT 13
BQ931374
LOCUS
DEFINITION BQ931374 915 bp mRNA linear EST 21-AUG-2002
AGENCOURT 8950013 NCI CGAP Mam2 Mus musculus cDNA clone
IMAGE:6442308 5', mRNA sequence.
ACCESSION BQ931374
VERSION BQ931374.1 GI:22346405
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 915)
AUTHORS NTH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13966 row: o column: 13
High quality sequence stop: 669.

FEATURES
source 1..915
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6442308"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam2"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 231 a 207 c 193 g 283 t 1 others
ORIGIN

Query Match 52.0%; Score 528.6; DB 13; Length 915;
Best Local Similarity 84.8%; Pred. No. 1.4e-140;
Matches 696; Conservative 0; Mismatches 94; Indels 31; Gaps 8;
QY 116 GTTCTGGGCTCAGGGGTGAGACCCCTCCGTAAGTCCGAGACCGGTGTGACAGACAG 175
Db 10 GGTTCGTGAGCCAGGGGTGAGACCCCTTCCAGAGTTCACAGACCGGTGTGACAGACAG 69
QY 176 CCTCCTTAATGACACGCTGCCATGTACGACCGCTTAACCTATACAGCCCATGCTCATAC 235
Db 70 CCTCCTTAATGACACGCTGCCATGTACGACCGCTTAACCTATACAGCCCATGCTCATAC 129
QY 236 GTAACCTTGTACTGTACGTAC---GATGGGTGTACAGCTCTGCTCTTTGATTTCATAG 292
Db 130 GTAACCTTGTACTGTACGTACGTGATGATGATGATGATGATGATGATGATGATGAT 189
QY 293 TGAGTCTCTTAAATACCACTGACCGGCTTTCGACGGCTTGAACAGAACTCTGGGTCC 352
Db 190 TGAGTCTCTTAAATACCACTGACCGGCTTTCGACGGCTTGAACAGAACTCTGGGTCC 249
QY 353 TGTGTTGCCCTTAACGAAGTATTCTGTCCTAGTCTGCTGGGTGCTGGGTGAGTGTG 412
Db 250 TGTGTTGCCCTTAATGAAGTATTCTGTCCTAGTCTGCTGGGTGCTGGGTGGA--GTGTG 307
QY 413 AAACACGACGCTCATCAAGGAGACAGACAGTATTTTGAATAATATGAAGTAGAGATTAAT 472
Db 308 AAACACGACATGATCAAGGAGACAGACAGTATTTTGAATAATATGAAGTAGAGATTAAT 367
QY 473 TTACACTATATTGTACATGAGTAATTCACCTGAATAAAGTGTACGGGTAAAGCTTTT 532
Db 368 TTACACTATATTGTACATGAGTAATTCACCTGAATAAAGTGTACGGGTAAAGCTTTT 427
QY 533 TAACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCGCATCTTATCAGTGCTC--- 589
Db 428 TAACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCGCATCTTATCAGTGCTCTTG 487
QY 590 -----TCTTGAGCCCTTCCCTCTGCTGCTCCCTCCAGATGGGCGGTGTAGTC 640
Db 488 AGCATCCCTTGGCGCCCCCACCCTCCACTGCGCTCCCTCCAGATGGGCGGTGTAGTC 547
QY 641 CATATTAAACTGGCCATCTCTCAGTGTCTAACTTAGCAAGTCTTTCTTTTAGGACCC 700
Db 548 CATATTAAACTGGCCATCTCTCAGTGTCTAACTTAGCAAGTCTTTCTTTTAGGACCC 605
QY 701 CCTTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATTAATGA----- 755
Db 606 CCTTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAACGAGGAGA 665
QY 756 --TTCGAGAGATTTTGTGAGTAGTAGTAGGTCGCTCTGCTTTTCCACCTTCTTTACT 813
Db 666 TTTTGTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTACT 725

QY 814 CAGCTG--ACTAGTCTCCCTTCGTTTCTAGTAACCTGGGTAGAAATCAGTCTGC 871
 |||
 Db 726 CAGAGCTAGTAGTCTCTTCTAGTTTCTAGSCAACTAAGTGTGCAATCATGTGTGC 785
 |||
 QY 872 GCGTTTACA---GTTTTTAACTATTTTATAGATTCTGAAA 909
 |||
 Db 786 CAGCTTTTACGGGTTTTTAACTATTTTATAGATTCTTTAA 826
 |||

RESULT 14
 AA817948/c
 LOCUS
 DEFINITION
 UI-R-A0-ag-e-05-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
 UI-R-A0-ag-e-05-0-UI 3' similar to dbj|D25274|HUMPO2ST9 Human
 randomly sequenced mRNA, mRNA sequence.
 AA817948
 AA817948.1 GI:4198471
 EST.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 545)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 889548
 On Feb 17, 1998 this sequence version replaced gi:2887828.
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEERF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dt track served to identify it as a clone from the normalized
 adult kidney library. cDNA Library Preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics This clone is also available through the I.M.A.G.E.
 Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1767351
 Seq primer: M13 Forward
 POLVA=No.

FEATURES

source
 1..545
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-A0-ag-e-05-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-A0"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; This library
 consists of a mixture of individually tagged normalized
 libraries constructed from rat placenta, adult lung, brain
 /liver, kidney, heart, spleen, ovary, and muscle. The tag
 is a string of 3-5 nucleotides present between the Not I
 site and the oligo-dt track which allows identification of
 the library of origin of a clone within the mixture."
 148 a 120 c 119 g 158 t

BASE COUNT
 ORIGIN
 Query Match 51.7%; Score 526; DB 9; Length 545;
 Best Local Similarity 100.0%; Pred. No. 5.7e-140;
 Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CCCCTATTCTGCTCAGATTAGAAATGCCAAATACCTTGTGACTAAGTTGCGTTGTG 60

Db 540 CCCTATTCTGCTCAGATTAGAAATGCCAAATACCTTGTGAACTAAGTTGCGTTGTG 481
 |||
 QY 61 CTGAGAACACGTAAGCACTAAGCTGTTGAGAGACTTTGTCCCTTAAGAAGACTCGACGTTTC 120
 |||
 Db 480 CTGAGAACACGTAAGCACTAAGCTGTTGAGAGACTTTGTCCCTTAAGAAGACTCGACGTTTC 421
 |||
 QY 121 TGGGCTCAGGGGTGCAGACCCCTCCCGTAGCTCCAGACCGTGTGACACAGCAGAGCTCC 180
 |||
 Db 420 TGGGCTCAGGGGTGCAGACCCCTCCCGTAGCTCCAGACCGTGTGACACAGCAGAGCTCC 361
 |||
 QY 181 TTAATGACACGCTGCCATGTAAGCACTTGTAACTTATCAGCCCATGCTCATTAACGTAAC 240
 |||
 Db 360 TTAATGACACGCTGCCATGTAAGCACTTGTAACTTATCAGCCCATGCTCATTAACGTAAC 301
 |||
 QY 241 TTGTACTCTAGCTCAGATGGGTGTAAACAGCTCTGCTCTTTGATTTCATAGAGTTCT 300
 |||
 Db 300 TTGTACTCTAGCTCAGATGGGTGTAAACAGCTCTGCTCTTTGATTTCATAGAGTTCT 241
 |||
 QY 301 CTAAATACACAGCTGACCGGCTTCTGCAGGCTTTGAACAGAACTCTGGCTCCCTGTGTGC 360
 |||
 Db 240 CTAAATACACAGCTGACCGGCTTCTGCAGGCTTTGAACAGAACTCTGGCTCCCTGTGTGC 181
 |||
 QY 361 CTCTAACGAAGTATTCTGTTCTAGTCTGGGTGTGCTGGGTGGAGTGTGTGAAACACGA 420
 |||
 Db 180 CTCTAACGAAGTATTCTGTTCTAGTCTGGGTGTGCTGGGTGGAGTGTGTGAAACACGA 121
 |||
 QY 421 CGTCATCAAGGAGACAGACAGTATTTTCACTAATATGAAGTAGAGATTAAATTTACACTA 480
 |||
 Db 120 CGTCATCAAGGAGACAGACAGTATTTTCACTAATATGAAGTAGAGATTAAATTTACACTA 61
 |||

RESULT 15

BQ201343/c
 LOCUS
 DEFINITION
 UI-R-DQ1-clt-n-08-0-UI.s1 UI-R-DQ1 Rattus norvegicus cDNA clone
 UI-R-DQ1-clt-n-08-0-UI 3', mRNA sequence.
 BQ201343
 BQ201343.1 GI:20417808
 EST.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 639)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 889548
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEERF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 The sequence contained an oligo-dt track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dt track served to verify it as a clone from the
 normalized rat cell line R3327-5a library cDNA library Preparation:
 M.B. Soares Lab Clone distribution: clones will be available
 through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLVA=yes.

FEATURES
Source

Location/Qualifiers

```
1. 639
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DQ1-clt-n-08-0-UI"
/dev_stage="adult"
/lab_host="DHI0B (Life Technologies)"
/clone_lib="UI-R-DQ1"
/note="Vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DQ1
library is a normalized Rat Cell Line R3327-5A library
(nRSA) constructed in pMT37 PAC vector according to the
procedure described by Bonaldo, Lennon & Soares
(Normalization and Subtraction: Two Approaches to
Facilitate Gene Discovery. Genome Research 6: 791-806,
1996). The oligonucleotide used to prime first strand
synthesis contained the sequence tag GGAAGTATC between
the Not I cloning site and dT18 stretch. The Rat Cell Line
R3327-5A was provided by Mary Hendrix of the University of
Iowa.
TAG_LIB=UI-R-DQ1
TAG_TISSUE=rat cell line R3327-5a
TAG_SEQ=GGACTAGATC"
BASE COUNT 183 a 130 c 136 g 130 t
ORIGIN

Query Match 51.7%; Score 526; DB 13; Length 639;
Best Local Similarity 100.0%; Pred. No. 6.3e-140;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATCTTGTGCAGATTAGAAATGCCAAATACCTTGTGAACCTAAGTTGGTTGTG 60
Db 541 CCCCTATCTTGTGCAGATTAGAAATGCCAAATACCTTGTGAACCTAAGTTGGTTGTG 482

QY 61 CTGAGAACAGCTAGACTAAGCTGTTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120
Db 481 CTGAGAACAGCTAGACTAAGCTGTTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 422

QY 121 TGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGGTGTCACACAGACGCTCC 180
Db 421 TGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGGTGTCACACAGACGCTCC 362

QY 181 TTAATGACAGCTGCCATGTAAAGCAGCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 240
Db 361 TTAATGACAGCTGCCATGTAAAGCAGCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 302

QY 241 TTTGTACTGTACGTACGATGGGTGTAAACAGCTCTGCTCTTTGATTTTATAGTGAGTTCT 300
Db 301 TTTGTACTGTACGTACGATGGGTGTAAACAGCTCTGCTCTTTGATTTTATAGTGAGTTCT 242

QY 301 CTAAATACAGCTGACCGGCTTCTGAGGCTTTGAAACAGAACTCTGCTCCTGTGTTC 360
Db 241 CTAAATACAGCTGACCGGCTTCTGAGGCTTTGAAACAGAACTCTGCTCCTGTGTTC 182

QY 361 CTCTACGAGTATCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTAACACGA 420
Db 181 CTCTACGAGTATCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTAACACGA 122

QY 421 CGTCATCAAGGAGACAGACAGTATTTTGAATATATGAAGTAGAGATTAAATTACACTA 480
Db 121 CGTCATCAAGGAGACAGACAGTATTTTGAATATATGAAGTAGAGATTAAATTACACTA 62

QY 481 CATGTACATGGAGTAAATCAACTGAATAAAAGTGTACCGGTAAA 526
Db 61 CATGTACATGGAGTAAATCAACTGAATAAAAGTGTACCGGTAAA 16
```

Search completed: November 23, 2003, 15:50:06
Job time : 2395.73 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:50:08 ; Search time 77.549 Seconds
(without alignments)
5788.425 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017
Sequence: 1 cccattcttctgctagatt.....cctttgggtctgagggttc 1017

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	187.2	18.4	605	3	US-09-385-982-483
C 2	68	6.7	588	3	US-09-385-982-133
C 3	54.2	5.3	7218	1	US-08-232-463-14
C 4	35.4	3.5	2038	1	US-08-181-271A-1
C 5	35.4	3.5	2038	1	US-08-449-515-1
C 6	35.4	3.5	2038	1	US-08-444-803-1
C 7	35.4	3.5	2038	1	US-08-449-043-1
C 8	35.4	3.5	2038	1	US-08-456-265A-1
C 9	35.4	3.5	2038	1	US-08-455-416-1
C 10	35.4	3.5	2038	1	US-08-455-244-1
C 11	35.4	3.5	2038	1	US-08-454-876-1
C 12	35.4	3.5	2038	2	US-08-457-364-1
C 13	35.4	3.5	2038	2	US-08-456-262-1
C 14	35.4	3.5	2038	2	US-08-456-240-1
C 15	35.4	3.5	2038	2	US-08-455-736-1
C 16	35.4	3.5	2038	2	US-08-455-736-1
C 17	35.4	3.5	2038	3	US-09-350-600-1
C 18	34	3.3	5084	1	US-08-306-691B-21
C 19	34	3.3	5084	5	PCT-US93-06251-25
C 20	33.2	3.3	15393	4	US-09-453-702B-191
C 21	32.2	3.2	453	4	US-09-484-970B-85
C 22	32	3.1	43360	4	US-09-453-702B-206
C 23	32	3.1	45325	4	US-09-453-702B-261
C 24	31.6	3.1	2703	4	US-09-482-273-75
C 25	31.6	3.1	2709	4	US-09-482-273-101
C 26	31.6	3.1	3001	4	US-09-539-333D-158
C 27	31.6	3.1	3001	4	US-09-539-333D-160

28	31.6	3.1	4203	2	US-08-866-757-1	Sequence 1, Appli
29	31.6	3.1	4203	3	US-09-153-593-1	Sequence 1, Appli
C 30	31.2	3.1	2559	4	US-09-569-098A-109	Sequence 109, App
31	31	3.0	1221	4	US-09-134-001C-365	Sequence 365, App
32	30.8	3.0	946	4	US-09-599-360B-42	Sequence 42, Appl
C 33	30.6	3.0	684	4	US-09-134-001C-1999	Sequence 1999, Ap
C 34	30.6	3.0	952	3	US-09-174-768-2	Sequence 2, Appli
35	30.6	3.0	58407	4	US-08-916-421B-2	Sequence 2, Appli
C 36	30.4	3.0	622	3	US-09-109-204-9	Sequence 9, Appli
C 37	30.4	3.0	622	4	US-09-490-032-9	Sequence 9, Appli
38	30.4	3.0	1422	4	US-09-227-357-18	Sequence 18, Appl
C 39	30.4	3.0	1975	3	US-09-109-204-4	Sequence 4, Appli
C 40	30.4	3.0	1975	4	US-09-490-032-4	Sequence 4, Appli
C 41	30.4	3.0	3010	4	US-09-996-243-313	Sequence 313, App
42	30.4	3.0	87350	3	US-08-781-891-79	Sequence 79, Appl
43	30.4	3.0	87350	4	US-09-618-166-79	Sequence 79, Appl
44	30.4	3.0	87543	4	US-09-791-211-3	Sequence 3, Appli
45	30.2	3.0	13865	3	US-09-009-217-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1

US-09-385-982-483/c
; Sequence 483, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CDDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-483

Query Match	18.4%	Score 187.2;	DB 3;	Length 605;
Best Local Similarity	65.5%;	Pred. No. 3.2e-52;		
Matches 348;	Conservative 0;	Mismatches 158;	Indels 25;	Gaps 5;
QY	485	GTACATGGAGTATTCAACTGAATAAAGTGCACGGTAAAGCTTTTAAACGGTTAAAT	544	
Db	597	GGACGAANGAANNAAGTCCCGGTAAGTAAAGTAAATTTTGTCAATNCACTA	538	
QY	545	TCGTCTCAACAGTAG-ATGACAAATGCCGATCTTATCAGTGTCTCTTGAGCCCCCT	603	
Db	537	GATAAANAANAAGTTTGTATACACAGNAATGTTTCTTANGCTTTTCTTNTTCTT	478	
QY	604	TCCCTCTGTCTCCCTCCCGATGGGGTTCAGTCCATATTTAAACTGGCCATCTCA	663	
Db	477	AACACTGCCATGCTCCCAATGSGGATTTAATCACTTAAACNGTGTCTGT	418	
QY	664	CAGTTCCTAACTAGCAAGTCTTTTCTTAGACCCCTTCTTAACGACCAATATCT	723	
Db	417	TAGTCGTAACCTTAGTAGTGTCTTTTCTTATAGAACCCCTTCTGACTGACCAATATG	358	
QY	724	GACCTGACTATAGATCTTCTGTATTAATGCATTCGAGATTTTGTGTAGATAGTA	783	

APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Utnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLSCULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-449-315-1
Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred.No.0.27; 96; Indels 2; Gaps 1;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GGCGTTGAGTCCATATTTAAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTGTCTTT 689
Db 314 GAGGTTAAATAACTATCAATTTGGACGTAAGAATATTTGGATATTTCTTCAGTTCTTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAAGGAGCAATATGTCGACCTGTACTATAAGATCTTCTG 747
Db 254 TCTAATTTCTTCCCTTAATTTCCACACITTGATGTTACTCATGTACTATCATATCTTCTA 195
QY 748 ATAATGCAATCGGAGATTTTTTTTGGTAGATAGTAGAAGTGCCTTCTCTGTTTTCACCTTC 807
Db 194 TTATCTCAITCCAATCAACGGCTTGATCTCGAGTATAAATAATTTCTTCTTCTTATCC 135
QY 808 TTTACTCAGCTGACTAGTCTTTC 830
Db 134 AATCAGCACTTGACITTAGATTC 112
RESULT 6
US-08-444-803-1/c
Sequence 1, Application US/08444803
Patent No. 5654414
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Utnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444.803
FILING DATE: 19-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-444-803-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

630 GGCGTTGAGTCCATATTTAAACTGGCCATCTCTCAGTTGCTAACTTAGCAAGTGTCTTT 689
631 GAGCTTAAATAAATATATCAAAATTTGGACGTAAGAAATATTTGGATATTTCTTCAGTTCTTTC 255
632 CTTTAGGACC--CCCTTCTTAAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTG 747
633 TCTAAATTTCTTCCCTTAAATTTCCAACTTGATGTTACTCATGTACTATCATCTTTCTA 195
634 ATAATGCATTTGGAGATTTTCTTGGTAGATAGTAGAAGTGGTTCCTGTGTTTTCACCTTCC 807
635 TTATCTCAATCCAAATCAGCGCTTGATCTCGAGTATAAATAAATTTCTATTTCTTATCCC 135
636 TTTACTCAGCTGACTAGTGCTTC 830
637 AATCAGCACTTGACTTTAGATTC 112

RESULT 7
US-08-449-043-1/c
Sequence 1, Application US/08449043
Patent No. 5689044
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,043
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/165,667
;; FILING DATE: 8-MAR-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/042,847
;; FILING DATE: 6-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/632,441
;; FILING DATE: 21-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/425,504
;; FILING DATE: 20-OCT-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,506
;; FILING DATE: 6-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/768,122
;; FILING DATE: 27-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/580,431
;; FILING DATE: 7-SEP-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/329,018
;; FILING DATE: 24-MAR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/045,957
;; FILING DATE: 12-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elmer, James Scott
;; REGISTRATION NUMBER: 36,129
;; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
;; TELEPHONE: (919)541-8614
;; TELEFAX: (919)541-8689
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2038 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 932..1435
;; US-08-449-043-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27; 96; Indels 2; Gaps 1;
Matches 105; Conservative 0; Mismatches 96

QY	630	GGCGTTAGTCCTATATTAACTGGCCATCCTCACAGTTGCTAACTAGCAAGTGCTTTT	689
Db	314	GAGCTTAATAACTATCAATTTGGACGTAAGAATATTGGATATCTTCAGTTCTTTC	255
QY	690	CTTTAGGAC--CCCTTCTTACGACCAATATGCTGACCTGTACTATAGATCTTTCTG	747
Db	254	TCTAATTTCTCTCCCTTAATTTCCAACTTGATGTTACTCATGTAATATCATATCTTCTA	195
QY	748	ATAATGCATTGGAGATTTTTTTGGTAGATAGAGAGCGGTCTCTGTTTTCACCTTCC	807
Db	194	TTATCTCATTCCAATCAGCGCTGATCTGGAGTATAAATAAATTTCTATTCTTAATCC	135
QY	808	TTTACTCAGTCACTAGTGCTTC	830
Db	134	AATCAGCACTGACTTTAGATTC	112

RESULT 8
US-08-456-265A-1/c
; Sequence 1, Application US/08456265A

;; Patent No. 5767369
;; GENERAL INFORMATION:
;; APPLICANT: Alexander, Danny C.
;; APPLICANT: Ryals, John A.
;; APPLICANT: Goodman, Robert M.
;; APPLICANT: Stinson, Jeffrey R.
;; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
;; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
;; NUMBER OF SEQUENCES: 111
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CIBA-GEIGY Corporation
;; STREET: 520 White Plains Road, P.O. Box 2005
;; CITY: Tarrytown
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10591
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/456,265A
;; FILING DATE: 31-MAY-95
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/181,271
;; FILING DATE: 13-JAN-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/093,301
;; FILING DATE: 16-JUL-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/937,197
;; FILING DATE: 6-NOV-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/678,378
;; FILING DATE: 1-APR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/305,566
;; FILING DATE: 6-FEB-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/165,667
;; FILING DATE: 8-MAR-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/042,847
;; FILING DATE: 6-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/632,441
;; FILING DATE: 21-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/425,504
;; FILING DATE: 20-OCT-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,506
;; FILING DATE: 6-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/768,122
;; FILING DATE: 27-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/580,431
;; FILING DATE: 7-SEP-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/368,672
;; FILING DATE: 20-JUN-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/329,018
;; FILING DATE: 24-MAR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/045,957
;; FILING DATE: 12-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meigs, J. Timothy
;; REGISTRATION NUMBER: 38,241

Db 314 GAGGTTAAATACTATCAATTTGGCGTAAGATATTTGGATATCTTCAGTTCTTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAAGAGCAATATGCTGACCTGTACTATAAGATCTTTCTG 747
Db 254 TCTAATTTCTTCCCTTAATTTCCAACTTGATGTTACTCATGTACTATCATATCTTTCTA 195
QY 748 ATAATGCATTCGGAGATTTTGGTAGATAGTAGAGTGCCTGCTCTGTTTTCACCTTCC 807
Db 194 TTATCTCATCAATCAACGGCTTGATCTGGAGTATAATAAATTTCTATTTCTTATCC 135
QY 808 TTACTCAGCTGACTAGTCTTTC 830
Db 134 AATCAGACTTGACTTTAGATTC 112

RESULT 10

US-08-455-244-1/c
; Sequence 1, Application US/08455244
; Patent No. 5789214

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/455,244

FILING DATE: 31-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/181,271

FILING DATE: 13-JAN-94

APPLICATION NUMBER: US 08/093,301

FILING DATE: 16-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/937,197

FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/678,378

FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/305,566

FILING DATE: 6-FEB-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/165,667

; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 932..1435
; US-08-455-244-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;

Best Local Similarity 51.7%; Pred. No. 0.27;

Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 630 GCGTTGAGTCCATATTTAACTGGCCATCCTCACAGTTGCTAACTAGCAAGTGTCTTT 689

Db 314 GAGGTTAAATACTATCAAAATTTGGACGTAAGAATATTTGGATATCTTCAGTTCTTTC 255

QY 690 CTTTAGGACC--CCCTTCTTAAGAGCAATATGCTGACCTGTACTATAAGATCTTTCTG 747

Db 254 TCTAATTTCTTCCCTTAATTTCCAACTTGATGTTACTCATGTACTATCATATCTTTCTA 195

QY 748 ATAATGCATTCGGAGATTTTGGTAGATAGTAGAGTGCCTGCTCTGTTTTCACCTTCC 807

Db 194 TTATCTCATCAATCAACGGCTTGATCTGGAGTATAATAAATTTCTATTTCTTATCC 135

QY 808 TTACTCAGCTGACTAGTCTTTC 830

Db 134 AATCAGACTTGACTTTAGATTC 112

RESULT 11

US-08-454-876-1/c

; Sequence 1, Application US/08454876

; Patent No. 5804693

; GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,876
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672

FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
SEQUENCE CHARACTERISTICS:
SEQUENCE FOR SEQ ID NO: 1:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-454-876-1
Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GGCGTTGAGTCCATATTTAACTGGCCATCTCTACAGTTGCTAGTACGAAAGTCTTTT 689
Db 314 GACGTTAAATACTATCAAAATTTGGACGTAAAGAAATTTGGATATTTCTTCAGTTCTTC 255
QY 690 CTTTAGGACC--CCCTCTTAAACGAGCAATATGTCGACCTGTACTATAGATCTTTCTG 747
Db 254 TCTAATTTCTCTCCCTTAATTTCCAACTTGTGATTTACTCATGTACTATCATCTTTCTA 195
QY 748 ATAATGCATTGGAGATTTTTTTGGTAGATAGTAGAAGTGGTCTCTGTTTTCACCTTCC 807
Db 194 TTATCTCATTCCAATCAGCGCTTGATCTGGAGTATATAATAATTCTATTTCTTATCCC 135
QY 808 TTTACTCAGCTGACTAGTGCTTC 830
Db 134 AATCAGCACTTGACTTTAGATTC 112
RESULT 12
US-08-457-364-1/c
Sequence 1, Application US/08457364
Patent No. 5847258
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-457-364-1
Query Match 3.5%; Score 35.4; DB 2; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GCGTTGAGTCCATATTTAACTGGCCATCCCTCACAGTGTCACTAGCAAGTGTCTTT 689
DB 314 GACGTTAAATAACTATCAAAATTTGGACGTAAGAATATTTGGATATCTTCACTTTTC 255
QY 690 CTTTAGGACC--CCCTTCTTACGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTG 747
DB 254 TCTAATTTCTTCCCTTAATTTCCACATTTGAATGTTACTCATGTACTATCATCTTTCTA 195
QY 748 ATAATGCATTCGGAGATTTTTTGGTAGATAGTAGAGTGCCTTCTGTGTTTTCACCTTC 807
DB 194 TTATCTCATTTCCAATCACGGTTGATCTGGAGTATAAATAAATTTCTATTTCTTATCCC 135
QY 808 TTTACTCAGCTGACTAGTGCTTC 830
DB 134 AATCAGACTTGACTTTAGATTC 112
RESULT 13
US-08-456-262-1/c
Sequence 1, Application US/08456262
Patent No. 5851766
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,262
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-456-262-1

Query Match 3.5%; Score 35.4; DB 2; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GGCGTTGAGTCCATATTAACTGGCCATCCTCACAGTTCTGTAACGTGCTTTT 689
Db 314 GAGTTAAATACTATCAAAATTTGGAGTAAAGAAATATTTGGATATTCCTCAGTCTTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAAAGAGCAATATGCTCACCTGTACTATAAGATCTTTCTG 747
Db 254 TCTAATTTCTCCCTTAATTTCCAATCTGATGTACTCAATGACTAATCAATCTTTCTA 195
QY 748 ATAATGCAATCGGAGATTTTTTTGGTAGATAGTAGAGTGCCTGCTTTCCTTTCACCTTCC 807
Db 194 TTATCTCATCCCAATCACCGGCTTGATCTGGAGTATAAATAAATTTCTATTTCTATCCC 135

QY 808 TTTTACTAGCTGACTAGTGTTC 830
Db 134 AATCAGCTTGACTTTAGATTC 112

RESULT 14

US-08-456-240-1/C
; Sequence 1, Application US/08456240
; Patent No. 5856154
; GENERAL INFORMATION:
; APPLICANT: Rvals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr, Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,240
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:

```

, APPLICATION NUMBER: US 07/848,506
, FILING DATE: 6-MAR-1992
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/768,122
, FILING DATE: 27-SEP-1991
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/580,431
, FILING DATE: 7-SEP-1990
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/368,672
, FILING DATE: 20-JUN-1989
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/329,018
, FILING DATE: 24-MAR-1989
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/045,957
, FILING DATE: 12-APR-1993
, ATTORNEY/AGENT INFORMATION:
, NAME: Elmer, James Scott
, REGISTRATION NUMBER: 36,129
, REFERENCE/DOCKET NUMBER: S-19825//
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (919)541-8614
, TELEFAX: (919)541-8889
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 2038 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, MOLECULE TYPE: cDNA
, FEATURE:
, NAME/KEY: CDS
, LOCATION: 932..1435
, US-08-456-240-1

```

	Query Match	3.5%;	Score 35.4;	DB 2;	Length 2038;
	Best local Similarity	51.7%;	Pred. No. 0.27;		
	Matches 105;	Conservative 0;	Mismatches 96;	Indels 2;	Gaps 1;
QY	630	GGCGTTGAGTCCATATTAAACTGGCCATCCTCAGATTGCTAACTTAGCAAGTCTTTT	689		
Db	314	GACGTTAAATAACTATCAAAATTTGGACGTAAGAATATTTGGATATCTTCAGATTCCTTC	255		
QY	690	CTTTAGGACC--CCGCTCTTAAACGAGCAATATGCTGACCTGTACTATAAGCACTCTTCG	747		
Db	254	TCRAATTTCCGTTCCCTTAATTTCCAACTTGAATGTTACTCATGTACTATCATACTTTCTA	195		
QY	748	ATAATGCATTCGGAGATTTTTTTTGGTAGATAGTAGAGTGGCTTCTCTTTTTCACCTTCC	807		
Db	194	TTATCTCATTCATCAACGGCTTCATCTGGAGTATAATAAATTTCTATTTTCTTATCC	135		

```

RESULT 15
US-08-455-736-1/c
; Sequence 1, Application US/08455736
; Patent No. 580328
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederic
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc

```

APPLICANT: Payne, George B.
 APPLICANT: Sperrison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Shericca C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/455,736
 FILING DATE: 31-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-1994
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8614

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 932..1435

US-08-455-736-1

Query Match

Best local Similarity 3.5%; Score 35.4; DB 2; Length 2038;

Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 630 GGCGTTGAGTCCATATTTAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTCTTTT 689

Db 314 GAGTTAAATACTATCCAAATTTGGACGTAAAGAAATTTGGATATTCTTCAGTCTTTC 255

QY 690 CTTTAGSACC--CCCTTCTTAAAGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTG 747

Db 254 TCTAATTTCTTCCCTTAATTTCCAACTTGATGTACTCATGTACTATCATCTTTCTA 195

QY 748 ATAATGCAITCGGAGATTTTGGTAGATAGTAGAGTGGTCCGTCTCGTTTACCTTCC 807

Db 194 TTATCTCAITCCAATCACGGCTTGATCTGGAGTATAAATAATTTCTATTCTTATCC 135

QY 808 TTTACTCAGCTGACTAGTGCTTC 830

Db 134 ATCAGACTTGACTTAGATTC 112

Search completed: November 23, 2003, 15:53:44

Job time : 80.549 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 14:02:08 ; Search time 360.418 Seconds
(without alignments)
9221.962 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccattctctgcagatt.....cctttgggtctgagggttc 1017

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT NEW PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	290.6	28.6	1644	14	US-10-153-668-48
2	290.6	28.6	1646	14	Sequence 48, Appl
3	290.6	28.6	2051	9	US-10-153-668-46
4	290.6	28.6	2941	14	Sequence 46, Appl
5	285.2	28.0	1232	10	US-10-198-846-9874
6	283.2	27.8	2313	14	Sequence 9874, App
7	195.2	19.2	594	14	Sequence 1632, App
8	193.2	19.0	602	10	US-10-066-543-2864
9	193.2	19.0	602	10	Sequence 12919, A
10	193.2	19.0	602	13	Sequence 2864, Ap
11	187.8	18.5	549	14	Sequence 1654, Ap
12	187.8	18.5	567	14	Sequence 1654, Ap
13	187.2	18.4	605	11	Sequence 2878, Ap
14	185.4	18.2	477	14	Sequence 2964, Ap
15	172.2	16.9	432	10	Sequence 483, App
16	164	16.1	2167	12	Sequence 3022, Ap
					Sequence 7612, Ap
					Sequence 500, App

17	147.4	14.5	409	14	US-10-066-543-13	Sequence 13, Appl
18	126.4	12.4	285	9	US-09-294-093B-506	Sequence 506, App
19	123.2	12.1	416	10	US-09-960-352-8781	Sequence 8781, Ap
20	108.8	10.7	293	14	US-10-066-543-3077	Sequence 3077, Ap
21	103.2	10.1	3740	11	US-09-764-891-9984	Sequence 9984, Ap
22	98.2	9.7	581	10	US-09-998-598-972	Sequence 972, App
23	91.6	9.0	565	10	US-09-998-598-296	Sequence 296, App
24	85.6	8.4	492	10	US-09-878-178-1503	Sequence 1503, Ap
25	85.6	8.4	492	13	US-10-046-935-1503	Sequence 1503, Ap
26	85.6	8.4	492	14	US-10-146-502-1503	Sequence 1503, Ap
27	82.2	8.1	398	11	US-09-918-995-34569	Sequence 34569, A
28	68	6.7	588	11	US-09-871-161-133	Sequence 133, App
29	62.2	6.1	426	10	US-09-960-352-7342	Sequence 7342, Ap
30	59	5.8	404	10	US-09-960-352-14206	Sequence 14206, A
31	57.6	5.7	1651	14	US-10-198-846-13019	Sequence 13019, A
32	52.4	5.2	162	10	US-09-920-300A-1549	Sequence 1549, Ap
33	52.4	5.2	162	12	US-10-099-926-1549	Sequence 1549, Ap
34	52.4	5.2	162	13	US-10-033-528-1549	Sequence 1549, Ap
35	40.8	4.0	289	9	US-09-923-876-6297	Sequence 6297, Ap
36	37.4	3.7	13712	12	US-10-311-455-1504	Sequence 1504, Ap
37	37.2	3.7	565	12	US-10-027-632-133365	Sequence 133365, A
38	37.2	3.7	565	13	US-10-027-632-133365	Sequence 133365, A
39	37	3.6	260209	12	US-10-025-966A-23	Sequence 23, Appl
40	37	3.6	260209	12	US-10-265-071-23	Sequence 23, Appl
41	36.5	3.6	1581	14	US-10-198-846-6480	Sequence 6480, Ap
42	36.4	3.6	725	12	US-10-027-632-11287	Sequence 11287, A
43	36.4	3.6	725	13	US-10-027-632-11287	Sequence 11287, A
44	36.2	3.6	968	12	US-10-027-632-121396	Sequence 121396, A
45	36.2	3.6	968	13	US-10-027-632-121396	Sequence 121396, A

ALIGNMENTS

RESULT 1

US-10-153-668-48
; Sequence 48, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Aki
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
US-10-153-668-48

Query Match 28.6%; Score 290.6; DB 14; Length 1644;
Best Local Similarity 64.5%; Pred. No. 5.5e-83;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

```
QY 1 CCCCTATTCTTGCTCAGATTAGAAATTCGCAAAATACCTTGGAACCTAGTTGC---GTT 57
Db 407 CCCCCATTCTTGTTTCAGATTAGAGTTGCAAAATACCTTCTGAACTACACTGCATTGTT 466
QY 58 GTGCTGAGAACACGTAAGCACTTAAGCTGTGTAGAGACTTTGTCCTTAAGAACACTGCAGC 117
Db 467 GTGCGGAGAACACCGGAGCACTGAACCTTTGCAAGACCTTTCGCTTTTGAGAAGACCGGTAGC 526
QY 118 TTCTGGGTCCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCGAGACCGGTGTGCACACA 169
Db 527 TTCTCAGTTAGAGGTGCAGACACTTGTCTCCTATGTAGTTCTCAGATGCGTAAGCA 586
QY 170 GCACAGCTCTTAATGACACCGCTGCATGTAACGCACTGTAACTTAATCAGCCCATGCT 229
Db 587 GAACAGCTCCCGAATGAAGCGTTGCCATTGAACCTCACCAGTGAGTTAGCAGACGTGTT 646
QY 230 CATTACGTAACCTTCTACTGTACGTCACGATGGGTGTAACAGCTCTGCTCTTTGATTTC 289
Db 647 CCGGACATAACATTGTACTGTATGTAATGGAGTGACGTAGCAGCTCAGTCTTTGGATCAGTC 706
QY 290 TAGTGAGTTCTCTAAATACCAAGCTGACCGGCTTCTGCAGGCTTTGAACAGAACTCTGCG 349
Db 707 TTTTGTGATTTTATAGCGAGTTTCTGCACGAGCTTTTGGGAGATTTTGAACAGAACTG-- 764
QY 350 TCTGTGTTGCCCTTAACGAAGTATTCTGTTCTCTAGTCGTGGGTGTGCTGGGTGAGTGT 409
Db 765 ----CTATTTCCTCTAAATGAAGAAATTTCTGTT--TAGCTGTGGGTGTGCGGGTGGGGTGT 818
QY 410 GTGAACACGAGCTCATCAAGAGGAGACACAGACTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 819 GT-----GTGATCAAGAGCAACAGACAGTATTTTGACAAATACGAAGTGGAG-- 866
QY 469 TAATTTACACTACATGTACATGGAGTAA--TTCAACTGAATAAAAGTGTCAACGGGTAA 525
Db 867 --ATTTACACTACATGTACAGGAATGAAGTGTCAACGGGTAAAACTCTAAAGGTTA 924
QY 526 ACCTTTTAAACGGTTAAATTTCTGTCTCAACAGTAGATGACAAATGGCCGATCTTATCAGTG 585
Db 925 ATTCTGTCAAAATGCAATGATGATGAAAGAAAGGTTGTTATTAATCAGGAAATGTTTTCT 984
QY 586 TCTCTCTTGAGCCCTTCCCTGCTGCTCCTCCGAGATGGGGCGTTGAGTCCATAT 645
Db 985 TAAGCTTTTCCCTTCTCTTACCTGCGCATGCTCCCAAATGGSCANTTAATTCATCT 1044
QY 646 TTAACTGGCCATCTCAGATGCTAATGCTTACCTTACCAAGTGTCTTTTCTTTAGGACCCCTTC 705
Db 1045 TTAACTGGTGTCTGTGTAGTCGCTAATCTTAGTAAGTGTCTTTCTTTATAGAACCCTTC 1104
QY 706 TTAAAGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATTAATGCTCGGAGATT 765
Db 1105 TGACTGAGCAATATGCT--CCTTGTTATTAATAATCTTTCTGATAATGCAATTAGAAGTT 1163
QY 766 TTTTGTGTAGTAGTAGAAGTGTCTCTGTTTTCACCTTCCTTTACTCAGCTGACTAGT 825
Db 1164 TTTTGTGCAATGATTAAGTGTCTTCCATGTAC-----TTTATTCAGAGCTAATAAGT 1218
QY 826 GTTTCCTCTGTTTCTTAGTAATCGGGTGTAGAAATCAGTGTGCTGGGCTTTACAGTTT 885
Db 1219 GCTTTCCTTAGTATTTCTAGTAACCTAGTGTAAAAATCATGTGTGTGACGCTTTATAGTTT 1278
QY 886 TAAACTATTTTAGATA-----TTCTGAAACATCACTCTCTTTCGCCAGAG 928
Db 1279 TAAATATTTTAGATAATTTTAACTATGAACCTTCTTAAACATCACTCTTTCGCCAGAT 1338
QY 929 TACCAACACTGTCTATGTGATGATGCGGCCCTCTAGACCTCACCACCGGAGACATG 988
Db 1339 TACCACACTGTCTACTTGACCAATCTG-ACCCTCTTTACCTGCCCCACCGGAGACAG 1397
QY 989 CTTCCGGTA 997
Db 1398 CTTCTCTGTA 1406
```

```
RESULT 2
US-10-153-668-46
; Sequence 46, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(273)
US-10-153-668-46
```

```
Query Match      28.6%; Score 290.6; DB 14; Length 1646;
Best Local Similarity 64.5%; Pred. No. 5.5e-83;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATTCTTGCTCAGATTAGAAATTCGCAAAATACCTTGGAACCTAGTTGC---GTT 57
Db 409 CCCCCATTCTTGTTTCAGATTAGAGTTGCAAAATACCTTCTGAACTACACTGCATTGTT 468
QY 58 GTGCTGAGAACACGTAAGCACTTAAGCTGTGTAGAGACTTTGTCCTTAAGAACACTGCAGC 117
Db 469 GTGCGGAGAACACCGGAGCACTGAACCTTTGCAAAAGACCTTTCGTCCTTTGAGAAGACGTTAGC 528
QY 118 TTCTGGGTCCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCGAGACCGGTGTGCACACA 169
Db 529 TTCTCAGTTAGAGGTGCAGACACTTGTCTCTCCTATGTAGTTCTCAGATGCGTAAGCA 588
QY 170 GCACAGCTCTCTTAATGACACGCTGCGCATGTAAACGACCTGTAACTTATCAGCCCATGCT 229
Db 589 GAACAGCTCTCCGAATGAAGCGTTGCCATTGAACCTCACCAGTGAGTTAGCAGACCGTGT 648
QY 230 CATTACGTAACCTTGTACTGTACGTCAAGTGGGTGAACAGCTCTGCTCTTTGATTTC 289
Db 649 CCGGACATAAATGTTACTGTAAATGGAGTGAGCGTAGCAGCTCTTTGGATCAGTC 708
QY 290 TAGTGAGTTCTCTAAATACCAAGCTGACCGGCTTCTGACGGCTTTGAAACAGAACTCTGGC 349
Db 709 TTTTGTGATTTTATAGCGAGTTTCTGACCGAGCTTTTTCGCGAGATTTTGAACAGAACTG-- 766
QY 350 TCTGTGTTGCCCTTAACGAAGTATTCTGTTCTCCTAGTCGTGGGTGTGCTGGGTGAGTGT 409
Db 767 ----CTATTTCCTCTAAATGAAGAAATTTCTGTT--TAGCTGTGGGTGTGCGGGTGGGGTGT 820
QY 410 GTGAACACGAGCTCATCAAGAGGAGACACAGACTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 821 GT-----GTGATCAAGAGCAAAAGACAGTATTTTGACAAATAACGAGTGGAG-- 868
```

```

QY 469 TAAATTTACACTACATGTGTACATGAGTAA---TTCAACTGAATAAAGTCTCAGGGTAA 525
Db 869 --ATTTACACTACATGTGTACAGAAATGAAGTGTACGGGTAAAGTCTTAAAGGTAA 926
QY 526 AGCTTTTAAACGGTAAATTTCTCTCAAAAGTAGATGACAAATGGCGAFCTTTATCAGTG 585
Db 927 ATTTCGTCAATGCAAGTAGATGATGAAGAAAGGTTGGTATTATCAGGAAATGTTTCT 986
QY 586 TCTCTCTTGAGCCCTTCCCTCGTGTCTCCCTCCAGATGGGGGTTGATGCCATAT 645
Db 987 TAACTGTTTCTCTCTTACACCTGCGATGCTCCCAAAATGGGCATTTAATTCATCT 1046
QY 646 TTAACCTGGCCATCTCACAGTTGCTAACTAGCAAGTCTTTCTTTAGACCCCTTC 705
Db 1047 TTAACCTGGTTGTTCTGTAGTGTCTAACTAGTAAAGTCTTTCTTTATGAACCCCTTC 1106
QY 706 TTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATAATGCAATTCGGAGATT 765
Db 1107 TGACTGAGCAATATGCT--CTTGTATTATAAATCTTTCTGATAATGCAATAGAGTT 1165
QY 766 TTTTGGTAGATAGTAGAGTGGTCTGCTTTTACCTTCCCTTACTCAGCTGACTAGT 825
Db 1166 TTTTGTGATTTAGTAAAGTGTCTTCCATGTTAC-----TTTATTACAGAGCTAATAAGT 1220
QY 826 GCTTCCCTTCTGTTTCTAGTAACTGGGTGTAGAAATCAAGTGTGCGGCTTTTACAGTTTT 885
Db 1221 GCTTCTCTAGTTTCTAGTAACTAGTGTAAAGTGTGTTGAGCTTTTATAGTTTT 1280
QY 886 TAAACTATTTTAGATA-----TTCTGAACATCACTGCTTGGCCAGAG 928
Db 1281 TAAATATTTTAGATAATCTTAAACTATGAACCTTCTTAACTCACTGCTTGGCCAGAT 1340
QY 929 TACCACACTGTGATGTATGATGCGGCCCTCTAGACCTCACCCGCGGACACATG 988
Db 1341 TACCACACTGTCACTTGACCAATCTG-ACCCTCTTTACCTGCGCCACCGGACACAG 1399
QY 989 CTTCCGGTA 997
Db 1400 CCTCCTGTA 1408

```

RESULT 3

```

US-09-925-302-255
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)

```

```

; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)
; OTHER INFORMATION: n equals a.t.g, or c
US-09-925-302-255

Query Match      28.6%; Score 290.6; DB 9; Length 2051;
Best Local Similarity 64.5%; Pred. No. 6.5e-83;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCTATTCTTGTCTCAGATTAAAGATTGCCAAATACCTTGTCAACTAAAGTTGC---GTT 57
Db 755 CCCCCCAATCTTGTTCAGATTAGAGTTTGCAAATACCTTCTGAACTTACACTGCAATGTT 814
QY 58 GTGCTGAGAACACGTAAGCACTTAAGCTGTTGTAGAGACTTTGTCTTAAAGAACTGCAGC 117
Db 815 GTGCCGAGAACACCGAGCACTGAACCTTTGCAAAAGACCTTCGTCTTTGAGAAGACGGTAGC 874
QY 118 TTTCTGGGCTCAGG-GGTGCAGACCCCTCCGTAGC-----TCCCAGACCGGTGACACA 169
Db 875 TTTCTGAGTTAGAGGTGACAGACACTTGTCTCTCTATGTAGTTCTCAGATGCGTAAGCA 934
QY 170 GCACAGCCCTCTTAAATGACACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCT 229
Db 935 GAACACCTCCCGAATGAAGCGTTGCCATTGAACTCACAGTGAAGTTAGCAGACAGTGT 994
QY 230 CATTACGTAACCTTTGACTGTACGTCACGATGGGTGTAAACAGCTCTGCTCTTTGATTTCA 289
Db 995 CCGGACATAAATGTTACTGTAAATGAGTGAGCGTAGCAGCTCAGCTCTTTGGATCAGTC 1054
QY 290 TAGTGTAGTTCTTAAATACCACTGACCGCTCTTCTGCAGGCTTTGAACAGAACTCTGGC 349
Db 1055 TTTGTGATTTCAATAGCGAGTTTCTGACGAGCTTTTTCGGGAGATTTTGAACAGAACTG-- 1112
QY 350 TCTGTGTGTCCTTAAACGAAATATTCTGTCTCTAGTCGTGGGTGTCTGGGTGGAGTGT 409
Db 1113 ----CTATTTCCTTAATGAAGAAATCTGTT--TAGCTGTGGGTGTGCGGGTGGGTGT 1166
QY 410 GTGAAACACAGACGTATCAAAAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 1167 GT-----GTGATCAAAGGACAAAGACAGATTTTGGACAAATACGAAAGTGGAG-- 1214
QY 469 TAAATTTACACTACATGTGTACATGGAGTAA---TTCAACTGAATAAAGTGTACGGGTAA 525
Db 1215 --ATTTACACTACATGTGTACAGGAATGAAGTGTCAACGGGTAAAGTCTCTAAAGGTTA 1272
QY 526 AGCTTTTAAACGGTAAATTTCTCTCAAAAGTAGATGACAAATGGCGATCTTTATCAGTG 585
Db 1273 ATTTCTGTCAATGCAAGTAGATGATGAAGAAAGTGGTATTATCAGGAAATGTTTCT 1332
QY 586 TCTCTCTTGAGCCCTTCCCTCGTGTCTCCCTCCAGATGGGGGTTGAGTCCATAT 645
Db 1333 TAAAGCTTTTCTTCTCTTACACCTGCCATGCTCCCAAAATTTGGCATTTAATTCATCT 1392
QY 646 TTAACCTGGCCATCTTCAAGTGTCTTAACTAGCAGTGTCTTTCTTTAGGACCCCTTC 705
Db 1393 TTAACCTGGTGTCTCTGTAGTGTCTTAACTAGTAAAGTCTTTTCTTTATAGAACCCCTTC 1452
QY 706 TTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATAATGCAATCGGAGATT 765
Db 1453 TGACTGAGCAATATGCT--CTTGTATTATAAATCTTTCTGATAATGCAATAGAGTT 1511
QY 766 TTTTGGTAGATAGTAGAAGTGGTTCCTGTTTCACTTCCCTTACTCAGCTGACTAGT 825
Db 1512 TTTTGTGATTAGTAAAGTGTCTTCCATGTTAC-----TTTATTACAGAGCTAATAAGT 1566
QY 826 GCTTCCCTTCTGTTTCTAGTAACTGGGTGTAGAAATCAGTGTGCTGGGCTTTACAGTTTT 885
Db 1567 GCTTCTCTTAGTTTCTAGTAACTAGGTAAAGTCAATGTGTGTGAGCTTTTATAGTTTT 1626
QY 886 TAAACTATTTTAGATA-----TTCTGAACATCACTGCTTGGCCAGAG 928
Db 1627 TAAATATTTTAGATAATCTTAAACTATGAACCTTCTTAACTCACTGTCTTTGCCAGAT 1686

```



```

QY 65 GAACACGTAAGCACTAAGCTGTTGAGAGACTTTGTCCTTAAGAGACTGCAGCTTCTGGG 124
Db 61 GAACACCGGACACTGAACCTTAGCAAGACCTTGTCTTTGAGAAGACGGTAGCTTGTGCA 120
QY 125 CTCAGG-GGTGCAGACCTCCCGTAGC-----TCCGAGACCGGTGTGACACAGCACAGC 176
Db 121 GTTAGGAGGTGCAGACACTTGTCTCTCCTATGATGTTCTCAGATCGGTAAAGCAAGACAGC 180
QY 177 CTCCTTAATGACAGCTGCCATGTAACGCACCTGTAACCTTATCAGCCCATGCTCATTAAG 236
Db 181 CTCGGAATGAAGCGTTGCCATTGAACCTCACAGTGAGTTAGCAGCAGCTTCCCGACA 240
QY 237 TAACCTTTGTAAGTACGTACAGTGGGTGTAACAGCTCTGCTTTTGNATTCATAGTAG 296
Db 241 TAACATTTGACTGTAAGGTAGTAGAGCTAGCTAGCTCTTTGGATCAGCTTTTGTGA 300
QY 297 TTCTCTAAATACACAGTGACCGGCTTTCTGAGGCTTTGAACAGAACTCTGGCTCCTGTG 356
Db 301 TTTTCATAGCGAGTTTCTGACCAAGCTTTTGGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGCTCTAAGAGTATTTCTGTTCTAGTCTGCTGGGTGCTGGGTGGAGTGTGAAAC 416
Db 355 TTTCTCTAATGAAGAAATCTGTGTT--TAGCTGTGGGTGCTGGCGGTGGGTGTGT----- 407
QY 417 ACGACGTCATCAAGAGGACAGACAGTATTTTGAC-TAATATGAAGTAGAGATTAATTTA 475
Db 408 -----GTGATCAAGAGCAAGACAGATTTTGAACAAATACGAAGTGGAG-----ATT 458
QY 476 CACTACATTTGATCATGAGT---AAATCAACTGAATAAAAGTGTACCGGTGAAGCTTTT 532
Db 459 CACTACATTTGATCAAGAGTAAAGTGTACCGGTGAAGAACTCTAAAGGTTAATTTCTG 518
QY 533 TAACGGTTAATTTCTGTCACACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTCT 592
Db 519 TCAATGCAATGATGATGAAGAAAGAGTGTGATTTATCAGGAATATGTTTCTTAAGCTT 578
QY 593 TGAGGCCCTTCCCTGCTGCTCCCTCCAGATGGGGGTGAGTCCATATTAAGT 652
Db 579 TTCTCTTCTTACACTGCTGCTCCCTCCCAATTTGGGCAATTAATCATCTTTAACT 638
QY 653 GGCATCTCACAAGTTGCTAACTTAGCAAGTGTCTTTTCTTAGACCCCTTCTTAACGA 712
Db 639 GGTGTGCTGTTAGTGTGCTAACTTAGTGTGCTTTCTTATAGAACCCCTTCTGACTGA 698
QY 713 GCAATATGCTGACCTGACTATAAGATCTTCTGATATGATGATGATGATGATGATGAT 772
Db 699 GCAATATGCTGCTTCTGATATTAATAATCTTCTGATATGATGATGATGATGATGATGAT 757
QY 773 TAGATAGTAGAGTGGTCTCTGTTTTCACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 832
Db 758 CGATTAGTAAAGTGTCTTCCATGTTAC-----TTTATTCAGAGCTAAATAGTGTCTTCC 812
QY 833 TTGCTTTCTAGTAAGTGGGTGAGAAATCACGTGCTGGGCTTTTACAGTTTAAACTA 892
Db 813 TTAGTTTCTAGTAAGTGGGTGAGAAATCATGTGTTGAGCTTTATAGTTTAAATA 872
QY 893 TTTTAGATA-----TTCTGAACATCTGCTTTCGCGAGATCAAC 935
Db 873 TTTTAGATAATTTTAAACTATGAACCTTCTTAAACATCTGCTTTCGCGAGATCAAC 932
QY 936 ACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995
Db 933 ACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 991
QY 996 TA 997
Db 992 TA 993

```

RESULT 6
 US-10-198-846-12919
 ; Sequence 12919, Application US/10198846

```

; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12919
; LENGTH: 2313
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-198-846-12919

```

```

Query Match      27.8%   Score 283.2; DB 14; Length 2313;
Best Local Similarity 63.5%; Pred. No. 1.8e-80;
Matches 651; Conservative 0; Mismatches 318; Indels 57; Gaps 12;

```

```

QY 1 CCCTATTTCTGCTCAGATTAAAGAAATTCGCAAAATACCTTGTGAACCTAAGTTGC---GTT 57
Db 1224 CCCCCATTTCTGCTCAGATTAAAGAAATTCGCAAAATACCTTGTGAACCTAAGTTGC 1283
QY 58 GTGCTGAGAACAGTAAGCACTAAGCTGTTGAGAGACTTTGCTTTAAGAGACTGCAGC 117
Db 1284 GTCCGAGAACACCGGACACTGAACCTTTGCAAAAGACCTTCGCTCTTGAAGACGCTAGC 1343
QY 118 TTCTGGGCTCAGG-GGTGACAGCCCTCCGTCAGC-----TCCGACACCGTGTGACACA 169
Db 1344 TTCTGAGTTAGAGGTGTCAGACACTTGTCTCTCTATGATGTTCTCAGATGCGTAAAGCA 1403
QY 170 GCACAGCCTCTTAAATGACACGCTGCCATGTAACGCACTGTAACTTATCAGCCCATGCT 229
Db 1404 GAACAGCCTCCCGAATGAAGCGTTGCCATGTAACCTCAGCTGAGTTAGCAGACGCTGTT 1463
QY 230 CATTAGTAACCTTTGACTGTAGCTCAGCATGGGTGTAAACAGCTGCTCTTTGATTTCA 289
Db 1464 CCGACATAAACAATGTAAGTGTATGAGTGTAGGAGTGTAGCAGCTCAGCTCTTTGATCAGTC 1523
QY 290 TAGTGAGTTCTCTAAATAATACCACTGACCGGCTTCTGAGGCTTTGACAGAACTCTGGC 349
Db 1524 TTGTGATTTTATAGCGAGTTTCTGACCAAGCTTTTGGGAGATTTTGAACAGAA----- 1578
QY 350 TCCTGTGTTGCTCTAACGAAGTATTTCTGTTCTCTAGTCTGGGTGTGCTGGGTGGAGTGT 409
Db 1579 --CTGTATTTCTCTAATGAAGAAATTTCTGTT--TAGCTTGTGTGTGCGGGTGGGTGT 1634
QY 410 GTGAACACAGCTCATCAAGAGAGACAGAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 1635 GT-----GTGATCAAGAGCAAAAGACAGTATTTTGACAAATAACGAAGTGTAGAT 1684
QY 469 TAATTTACACTCAATGTATGAGTGAATTTCACTGAATAAAGTGTACCGGTAAAGC 528
Db 1685 TACACTACATTTGACAGGAATGAA--AGTGTACGGGTAAAGAACTCTAAGAGTTAAT 1742
QY 529 TTTTAAACGGTTAATTTCTGTCAAAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCT 588
Db 1743 TCTGTCAATGACAGTAGATGATGAAGAAAGTGTGATATATCAGGAATGTTTCTTAA 1802
QY 589 CTCCTTGAAGCCCCCTTCCCTGCTGCTCCCTCCAGATGGGGCTGTAGTCCATATTTA 648
Db 1803 GCTTTTCTCTTTTACACTGCTCCCTCCCAATTTGGGCAATTTAATCACTTTA 1862
QY 649 AACTGGCATCTCAGCTGCTAAGTGTAGTGTCTTTTCTTTAGGACCCCTTCTTA 708
Db 1863 AACTGGTGTCTGTGTAGTGTAGTGTCTTTTCTTTATAGAACCCCTTCTGA 1922

```

QY 709 ACGAGCAATATGCTGACCTGCTACTATTAAGATCTTTCTGATAAGCAATTCGGAGATTTT 768
Db 1923 CTGAGCAATATGCTCT-CCTTGATTAATAAATCTTTCTGATAAGCAATTAAGAGGTTTTT 1981
QY 769 TTGGTAGATAGTGAAGTGGCTCTGTTTTCACCTCTCTTACTCAGCTGACTAGTGCT 828
Db 1982 TTGTGATTAGTAAAGTGGCTTTCATGTTAC-----TTTATTGAGAGCTAATAAGTGCT 2036
QY 829 TCCCTTCCTTTCTAGTAAGTGGGTGTAAGAAATCACGTGCTGCGCTTTTACAGTTTTTAA 888
Db 2037 TTCCTTAGTTCTAGTAACTAGGTGTAAGAAATCATGTTGTCAGCTTTATAGTTTTTAA 2096
QY 889 ACTATTTTAGATA-----TTCTGAACATACACTGTCTTGCACAGTAC 931
Db 2097 AATATTTTAGATAATCTTAAACTATGAACCTTTTAAATCACTGTCTTGCAGGTTAC 2156
QY 932 CAACACTGCTGATGATGATGCGCCCTCTAGACCTCACCCACGGGACACATGCTT 991
Db 2157 CGACACTGTCTTGAACCAATACG-ACCCTTTTACCTCGCCACGGGACACAGGCT 2215
QY 992 CCGGTA 997
Db 2216 CCGTA 2221

RESULT 7
US-10-066-543-2864
; Sequence 2864, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2864
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2864

Query Match 19.2%; Score 195.2; DB 14; Length 594;
Best Local Similarity 70.9%; Pred. No. 2.7e-52;
Matches 373; Conservative 0; Mismatches 113; Indels 40; Gaps 7;
QY 505 GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAAA--CAGTAGATG 562
Db 6 GAATGAAGTGTACGGGTAAAACTCTAAAGGTTAAATTTCTGTCAATGCAAGTAGATG 65
QY 563 ACAAA-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCCC 608
Db 66 ATGAAGAAAGGTTGGTATTATATCAGGAATGTTTCTTAAGCTTTTCTTCTTCTACAC 125
QY 609 CTGCTGTCCCTCCCGATGGGGTGTAGTCATATTTAAACTGGCCATCTTCACAGTT 668
Db 126 CTGCCATGCTTCCCAAAATTTGGGCAATTAATCATCTTTAAACTGGTGTCTGTAGTC 185
QY 669 GCTAACTTAGCAAGTGTCTTTTCTTTAGACCCCTCTTAAACGAGCAATATGCTGACCT 728

Db 186 GCTAACTTAGTAAGTGTCTTTTCTTATAGAACCCCTTCTGACTGACCAATATGCTT-CCTT 244
QY 729 GTACTATAAGATCTTTCTGTATAAGTCATTTCGGAGATTTTTTTGGTAGATAGTAGAGTGC 788
Db 245 GTATTATAAATCTTTCTGATAAGCAATTAAGGTTTTTTTGTGATTAGTAAAGTGC 304
QY 789 GTTCTGTGTTTTCACCTTCTTACTCAGCTGACTAGTCTTCCCTTCGTTTCTAGTAAC 848
Db 305 TTTCCATGTTAC-----TTTATTGAGAGCTAATAAGTGTCTTCTTCTAGTTTCTAGTAAC 359
QY 849 TGGGTGTAGAAATCACGTGCTGCGCTTTACAGTTTTTAAACTATTTTAGATA----- 901
Db 360 TAGGTGTAAAATCATGTGTTGTCAGCTTTATAGTTTTTAAAATATTTTAGATAATTCTTA 419
QY 902 -----TTCTGAACATACACTGTCTTGCACAGTACCAACACTGTCTGATGATGA 951
Db 420 AACTATGAACCTTCTTAAATCACTGTCTTGCAGATTACCGACACTGTCACTTGACCAA 479
QY 952 TGCCGCCCTCTCTAGACCTCACCCACGGGACACATGCTTCCGTA 997
Db 480 TACTG-ACCCTTTTACCTCGCCACGGGACACACGCTCCTGTA 524

RESULT 8
US-09-878-178-1654/c
; Sequence 1654, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(602)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1654

Query Match 19.0%; Score 193.2; DB 10; Length 602;
Best Local Similarity 70.5%; Pred. No. 1.2e-51;
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
QY 505 GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAAA--CAGTAGATG 562
Db 592 GAATGAAGTGTACGGGTAAANCTTAAAGGTTAAATTTCTGTCAATGCAAGTAGATG 533
QY 563 ACAAA-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCCC 608
Db 532 ATGAAGAAAGGTTGGTATTATCAGGAATGTTTCTTAAAGCTTTTCTTCTTCTACAC 473
QY 609 CTGCTGTCCCTCCCGATGGGGTGTAGTCATATTTAAACTGGCCATCTTCACAGTT 668
Db 472 CTGCCATGCTTCCCAAAATTTGGGCAATTAATCATCTTTAAACTGGTGTCTGTAGTC 413
QY 669 GCTAACTTAGCAAGTGTCTTTTCTTTAGACCCCTCTTAAACGAGCAATATGCTGACCT 728
Db 412 GCTAACTTAGTAAGTGTCTTTCTTATAGAACCCCTTCTGACTGAGCAATGCTT-CCTT 354
QY 729 GTACTATAAGATCTTTCTGATAATGCAATTCGGAGATTTTTTTGGTAGATAGTAGAGTGC 788
Db 353 GTATTATAAATCTTTCTGATAATGCAATTAAGGTTTTTTTGTGCGATTAGTAAAGTGC 294
QY 789 GTTCTGTGTTTTCACCTTCTTTTACTAGTGTACTAGTGTCTTCCCTTCTGTTTCTAGTAAC 848

Db 293 TTTCAGTTAC-----TTTATTAGAGCTAATAGTCTTCTTAGTTTTCAGTAAC 239
QY 849 TGGGTAGAAATCAGTGTGGCGTTTACAGTTTTTAACTATTTAGATA----- 901
Db 238 TAGGTGTAAAAATCATGTGTGACCTTTATAGTTTTTAAATATTTAGATAATTTCTTA 179
QY 902 -----TTCTGAACATCACTGTCTGCCAGAGTACCAACACTGTCAATGTATGA 951
Db 178 AACTATGAACCTTTTAACTATCACTGTCTGCCAGATTTAGCGACACTGGCACTTGACCAA 119
QY 952 TGGCGCCCTCTAGACCTCACCACCGGACACATGCTTCCGGTA 997
Db 118 TACTG-ACCCTTTTACTCGCCACCGGACACACGCTCTCTGTA 74

RESULT 9

US-10-046-935-1654/c
; Sequence 1654, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 569, 578
; OTHER INFORMATION: n = A,T,C or G

US-10-046-935-1654

Query Match 19.0%; Score 193.2; DB 13; Length 602;
Best Local Similarity 70.5%; Pred. No. 1.2e-51;
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
QY 505 GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 562
Db 592 GAATGAAGTGTACGGGTAAAGCTTTAAAGGTTAAATTTCTGTCAATGCGAGTAGATG 533
QY 563 ACAAAT-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCCC 608
Db 532 ATGAAGAAGTGTGTTGTTATATCAGGAATGTTTTCTTAAGCTTTTCTTCTTTACAC 473
QY 609 CTGCTCTCCCTCCCGAGATGGGCGTTGAGTCCATATTTAAACTGGCCATCTTCAAGTT 668
Db 472 CTGCCAIGCTCCCCAAATGGGCAATTAATCACTTTAAACTGGTTGTTCTGTGTAGTC 413
QY 669 GCTAATTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAAACGAGCAATATGTCACCT 728
Db 412 GCTAATTAGTAGTGTCTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CCCT 354
QY 729 GTACTATAGATCTTTCTGATATGCAATCGGAGATTTTTTGGTAGATAGTAGAGTGC 788
Db 353 GTATTATAAAATCTTTCTGATATGCAATAGAGGTTTTTTTGTGCGATTAGTAAAGTGC 294
QY 789 GTTCCTGTTTTACCTTCTTACTCAGCTGACTAGTGTCTTCCCTTCTTCTTAGTAAC 848
Db 293 TTTCAGTTAC-----TTTATTAGAGCTAATAGTGTCTTCTTAGTTTCTAGTAAC 239
QY 849 TGGGTAGAAATCAGTGTGGCGTTTACAGTTTTTAACTATTTAGATA----- 901
Db 293 TTTCAGTTAC-----TTTATTAGAGCTAATAGTGTCTTCTTAGTTTCTAGTAAC 239
QY 849 TGGGTAGAAATCAGTGTGGCGTTTACAGTTTTTAACTATTTAGATA----- 901

Db 238 TAGGTGTAAAAATCATGTGTGACCTTTATAGTTTTTAAATATTTAGATAATTTCTTA 179
QY 902 -----TTCTGAACATCACTGTCTGCCAGAGTACCAACACTGTCAATGTATGA 951
Db 178 AACTATGAACCTTTTAACTATCACTGTCTGCCAGATTTAGCGACACTGGCACTTGACCAA 119
QY 952 TGGCGCCCTCTAGACCTCACCACCGGACACATGCTTCCGGTA 997
Db 118 TACTG-ACCCTTTTACTCGCCACCGGACACACGCTCTCTGTA 74

RESULT 10

US-10-146-502-1654/c
; Sequence 1654, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 569, 578
; OTHER INFORMATION: n = A,T,C or G

US-10-146-502-1654

Query Match 19.0%; Score 193.2; DB 14; Length 602;
Best Local Similarity 70.5%; Pred. No. 1.2e-51;
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
QY 505 GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 562
Db 592 GAATGAAGTGTACGGGTAAAGCTTTAAAGGTTAAATTTCTGTCAATGCGAGTAGATG 533
QY 563 ACAAAT-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCCC 608
Db 532 ATGAAGAAGTGTGTTGTTATATCAGGAATGTTTTCTTAAGCTTTTCTTCTTTACAC 473
QY 609 CTGCTCTCCCTCCCGAGATGGGCGTTGAGTCCATATTTAAACTGGCCATCTTCAAGTT 668
Db 472 CTGCCAIGCTCCCCAAATGGGCAATTAATCACTTTAAACTGGTTGTTCTGTGTAGTC 413
QY 669 GCTAATTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAAACGAGCAATATGTCACCT 728
Db 412 GCTAATTAGTAGTGTCTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CCCT 354
QY 729 GTACTATAGATCTTTCTGATATGCAATCGGAGATTTTTTGGTAGATAGTAGAGTGC 788
Db 353 GTATTATAAAATCTTTCTGATATGCAATAGAGGTTTTTTTGTGCGATTAGTAAAGTGC 294
QY 789 GTTCCTGTTTTACCTTCTTACTCAGCTGACTAGTGTCTTCCCTTCTTCTTAGTAAC 848
Db 293 TTTCAGTTAC-----TTTATTAGAGCTAATAGTGTCTTCTTAGTTTCTAGTAAC 239
QY 849 TGGGTAGAAATCAGTGTGGCGTTTACAGTTTTTAACTATTTAGATA----- 901
Db 238 TAGGTGTAAAAATCATGTGTGACCTTTATAGTTTTTAAATATTTAGATAATTTCTTA 179
QY 902 -----TTCTGAACATCACTGTCTTCCAGAGTACCAACACTGTCAATGTATGA 951
Db 178 AACTATGAACCTTTTAACTATCACTGTCTTCCAGGATTTAGCGACACTGGCACTTGACCAA 119

QY 952 TGCGCCCGCTTAGACCTCACCACGGGACACATGCTTCGGTA 997
Db 118 TACTG-ACCCTCTTTTACCTCGCCACGGGACACACGCTCTCTGTA 74

RESULT 11

US-10-066-543-2878/c
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:

; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563

; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31

; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878

; LENGTH: 549

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 526

; OTHER INFORMATION: n = A,T,C or G

US-10-066-543-2878

Query Match 18.5%; Score 187.8; DB 14; Length 549;
Best Local Similarity 73.3%; Pred. No. 6.4e-50;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

QY 600 CCCTTCCCGCTGTGCTCCCTCCCGATGGGGTTGAGTCCCATTTTAACTGGCCATC 659
Db 481 CTCCTACCTGCGATGCTCCCTCCCAATTTGGGCAITTAATCATCTTTAACTGGTTGT 422
QY 660 CTCACAGTTGCTAACTAGCAAGTGTCTTTCTTAGGACCCCTTCTTAACGAGCAATAT 719
Db 421 CTGTTAGTCGTAACCTAGTAAGTGTCTTTCTTATAGAACCCCTCTGACTGAGCAATAT 362
QY 720 GTCTGACCTGTACTATAAGATCTTTCTGTATATGCAATCGGAGATTTTTTGGTAGATAG 779
Db 361 GCCT-CCCTGTATTATAAATCTTTCTGTATATGCAATCGGAGATTTTTTGGTAGATAG 303
QY 780 TAGAAGTGGTTCCTGTTTCCCTTTACCTTCTAGCTGACCTAGTGTCTCCCTTCGTTT 839
Db 302 TAAAGTGTCTTCCATGTTAC-----TTTATTACAGAGCTAATAAGTGTCTTCTTGTAGTT 248
QY 840 TCTAGTAAGTGGGTGTAGAAATCACTGTCTGGGGTTTACAGTTTTTAACTATTTTGA 899
Db 247 TCTAGTAAGTGTAGTGTAAATAATCATGTGTGACCTTTATAGTTTTTAAATAATTTTGA 188
QY 900 TA-----TTCTGAACATCACTGTCTGCCAGAGTACCAACACATGCA 942
Db 187 TAATCTTAAATATGAACCTTCTTAACATCATCTGTCTGCCAGATTAACGACATGCA 128
QY 943 TGTGATTGATGCGGCCCTCTAGACCTCACCACGGGACACATGCTTCCGGTA 997
Db 127 CTTGACCAATACTG-ACCCTCTTTACCTCGCCACGGGACACACGCTCTCTGTA 74

RESULT 12

US-10-066-543-2964/c
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:

; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563

; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31

; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2964

; LENGTH: 567

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 527

; OTHER INFORMATION: n = A,T,C or G

US-10-066-543-2964

Query Match 18.5%; Score 187.8; DB 14; Length 567;
Best Local Similarity 73.3%; Pred. No. 6.6e-50;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

QY 600 CCCTTCCCGCTGTGCTCCCTCCCGATGGGGTTGAGTCCCATTTTAACTGGCCATC 659
Db 481 CTCCTACCTGCGATGCTCCCTCCCAATTTGGGCAITTAATCATCTTTAACTGGTTGT 422
QY 660 CTCACAGTTGCTAACTAGCAAGTGTCTTTCTTAGGACCCCTTCTTAACGAGCAATAT 719
Db 421 CTGTTAGTCGTAACCTAGTAAGTGTCTTTCTTATAGAACCCCTCTGACTGAGCAATAT 362
QY 720 GTCTGACCTGTACTATAAGATCTTTCTGTATATGCAATCGGAGATTTTTTGGTAGATAG 779
Db 361 GCCT-CCCTGTATTATAAATCTTTCTGTATATGCAATCGGAGATTTTTTGGTAGATAG 303
QY 780 TAGAAGTGGTTCCTGTTTCCCTTTACCTTCTAGCTGACCTAGTGTCTCCCTTCGTTT 839
Db 302 TAAAGTGTCTTCCATGTTAC-----TTTATTACAGAGCTAATAAGTGTCTTCTTGTAGTT 248
QY 840 TCTAGTAAGTGGGTGTAGAAATCACTGTCTGGGGTTTACAGTTTTTAACTATTTTGA 899
Db 247 TCTAGTAAGTGTAGTGTAAATAATCATGTGTGACCTTTATAGTTTTTAAATAATTTTGA 188
QY 900 TA-----TTCTGAACATCACTGTCTGCCAGAGTACCAACACATGCA 942
Db 187 TAATCTTAAATATGAACCTTCTTAACATCATCTGTCTGCCAGATTAACGACATGCA 128
QY 943 TGTGATTGATGCGGCCCTCTAGACCTCACCACGGGACACATGCTTCCGGTA 997
Db 127 CTTGACCAATACTG-ACCCTCTTTACCTCGCCACGGGACACACGCTCTCTGTA 74

RESULT 13

US-09-871-161-483/c
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:

; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; PRODUCTS: II

FILE REFERENCE: CDNA-260XX
CURRENT APPLICATION NUMBER: US/09/871,161
CURRENT FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 09/328,111
PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/117,393
PRIOR FILING DATE: 1999-01-27
PRIOR APPLICATION NUMBER: 60/098,639
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 483
LENGTH: 605
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(605)
OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

Query Match 18.4%; Score 187.2; DB 11; Length 605;
Best Local Similarity 65.5%; Pred. No. 1.1e-49;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;
QY 485 GTACATGGAGTAATTCAACTGAATAAAGTGTCAAGGTAAGCTTTTAAACGGTTAATT 544
Db 597 GGCAANGAANAAGTCCGGGTAAACCTTTAAANGGTTAATTTTGTCAAATNCAGTA 538
QY 545 TCTGTCAACAGTAG-ATGCAAAATGGCCGATCTTATCAGTGTCTCTTTGAGCCGCCCT 603
Db 537 GATAAANAANAAGTTTGNATATAACAGNAATGTTTCTTANGCTTTTCNTTNTCTT 478
QY 604 TCCCTCTGCTGCTCCCTCCAGATGGGGCGTTAGTCCATATTTAAACTGCCCATCTCA 663
Db 477 AACACCTGCCATGCTCCCAAAATGGGCAATTTAAATCACTTTAAACNGTTGTTCTGT 418
QY 664 CAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTTCTTAACGACCAATATGCT 723
Db 417 TAGTCGCTAACTTAGTAAGTGTCTTTCTATAGAACCCCTTCTGACTGAGCAATATGCT 358
QY 724 GACCTGCTATTAAGATCTTTCTGATAATGCAATTCGGAGATTTTGTGGTAGATAGTA 783
Db 357 -CCTTGATTAATAAATCTTCTGATAATGCAATTCAGAGTCTTTTGTTCGANTAGTAA 299
QY 784 AGTGGCTTCTGTTTCACTTCCCTTTACAGTGTAGTCTTCCCTTCCCTTCCCTTCTA 843
Db 298 AGTGGCTTCTGTTTCACTTCCCTTTACAGTGTAGTCTTCCCTTCCCTTCCCTTCTA 244
QY 844 GTAACCTGGTGTAGAATCACTGCTGCGGCTTTACAGTGTAGTCTTCCCTTCCCTTCTA 901
Db 243 GTAACCTGGTGTAGAATCACTGCTGCGGCTTTACAGTGTAGTCTTCCCTTCCCTTCTA 184
QY 902 -----TTCTGAAACATCACTGTCTTGCAGAGTACCAACACTGTCATGTG 946
Db 183 TCTTAAACTATGAACCTTCTTAACATCACTGCTTGCAGATTAACGACACTGTCATGT 124
QY 947 ATTGATGCGCCCTCTAGACCTTACCACCGGACACATGCTTCCGGTA 997
Db 123 ACCAATAGT-ACCCTCTTTACTCGCCACCGGACACACGCTCTCTGTA 74

RESULT 14

US-10-066-543-3022/c
Sequence 3022, Application US/10066543
Publication No. US20030087818A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yuqiu
APPLICANT: Pyle, Ruth A.
APPLICANT: Xu, Jiangchun
APPLICANT: Indrias, Carol Yoseph
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather

APPLICANT: Carter, Darrick
APPLICANT: Fanger, Gary R.
APPLICANT: Smith, Carole L.
APPLICANT: Durham, Margarita
APPLICANT: Stolk, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.563
CURRENT APPLICATION NUMBER: US/10/066,543
CURRENT FILING DATE: 2002-01-31
NUMBER OF SEQ ID NOS: 3417
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3022
LENGTH: 477
TYPE: DNA
ORGANISM: Homo sapiens
US-10-066-543-3022

Query Match 18.2%; Score 185.4; DB 14; Length 477;
Best Local Similarity 73.2%; Pred. No. 3.5e-49;
Matches 301; Conservative 0; Mismatches 86; Indels 24; Gaps 4;
QY 604 TCCCTCTGCTGCTCCCTCCAGATGGGGCGTTAGTCCATATTTAAACTGCCCATCTCA 663
Db 477 TACACCTGCCATGCTCCCAAAATGGGCATTTAAATCATCTTTAAACTGGTGTCTGT 418
QY 664 CAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTTCTTAACGACCAATATGCT 723
Db 417 TAGTCGCTAACTTAGTAAGTGTCTTTCTATAGAACCCCTTCTGACTGAGCAATATGCT 358
QY 724 GACCTGCTATTAAGATCTTTCTGATAATGCAATTCGGAGATTTTGTGGTAGATAGTA 783
Db 357 -CCTTGATTAATAAATCTTCTGATAATGCAATTCAGAGTCTTTTGTTCGANTAGTAA 299
QY 784 AGTGGCTTCTGTTTCACTTCCCTTTACAGTGTAGTCTTCCCTTCCCTTCCCTTCTA 843
Db 298 AGTGGCTTCTGTTTCACTTCCCTTTACAGTGTAGTCTTCCCTTCCCTTCCCTTCTA 244
QY 844 GTAACCTGGTGTAGAATCACTGCTGCGGCTTTACAGTGTAGTCTTCCCTTCCCTTCTA 901
Db 243 GTAACCTGGTGTAGAATCACTGCTGCGGCTTTACAGTGTAGTCTTCCCTTCCCTTCTA 184
QY 902 -----TTCTGAAACATCACTGTCTTGCAGAGTACCAACACTGTCATGTG 946
Db 183 TCTTAAACTATGAACCTTCTTAACATCACTGCTTGCAGATTAACGACACTGTCATGT 124
QY 947 ATTGATGCGCCCTCTAGACCTTACCACCGGACACATGCTTCCGGTA 997
Db 123 ACCAATAGT-ACCCTCTTTACTCGCCACCGGACACACGCTCTCTGTA 74

RESULT 15

US-09-960-352-7612
Sequence 7612, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 7612
LENGTH: 432
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 33-LIB188-020-Q1-E1-A2
US-09-960-352-7612

Search completed: November 23, 2003, 19:05:51
Job time : 365.418 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:17:01 ; Search time 99.1939 Seconds
(without alignments)
9470.385 Million cell updates/sec

```

Title:          US-09-717-321A-17
Perfect score:  348
Sequence: 1 tgaacatcactctcttgc.....cttgaacttcaaaaaaaa 348

```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Genseq_19Jun03.*

1:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1980.DAT.*
2:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1981.DAT.*
3:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1982.DAT.*
4:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1983.DAT.*
5:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1984.DAT.*
6:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1985.DAT.*
7:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1986.DAT.*
8:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1987.DAT.*
9:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1988.DAT.*
10:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1989.DAT.*
11:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1990.DAT.*
12:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1991.DAT.*
13:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1992.DAT.*
14:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1993.DAT.*
15:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1994.DAT.*
16:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1995.DAT.*
17:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1996.DAT.*
18:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1997.DAT.*
19:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1998.DAT.*
20:	SIDS1/cgcdata/genseq/genseqn-emb1/NA1999.DAT.*
21:	SIDS1/cgcdata/genseq/genseqn-emb1/NA2000.DAT.*
22:	SIDS1/cgcdata/genseq/genseqn-emb1/NA2001A.DAT.*
23:	SIDS1/cgcdata/genseq/genseqn-emb1/NA2001B.DAT.*
24:	SIDS1/cgcdata/genseq/genseqn-emb1/NA2002.DAT.*
25:	SIDS1/cgcdata/genseq/genseqn-emb1/NA2003.DAT.*

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	348	100.0	348	22	AAH22399	Human rac1 gene re
2	348	100.0	1266	22	AAH22396	Human rac1 contig
3	209.4	60.2	2051	21	AAH18236	Lung cancer associ
4	203.6	58.5	28567	25	AB417030	Human M21 gene Ra
5	187.8	54.0	352	22	AAH22400	Human rac1 gene re
6	187.4	53.9	1232	24	ABR836623	Human cDNA differe
7	187.4	53.9	1232	24	AAH95134	Gene #1632 used to
8	187.4	53.9	1318	23	AA565569	DNA encoding novel

C	9	163.6	47.0	353	25	ABX46267	Bovine EST associ
	10	160.6	46.1	447	21	AAA89693	Mouse Exo103 nucle
	11	145.2	41.7	422	21	AAA89694	Mouse Rab2 nucleot
C	12	123	35.3	123	22	AAH22395	Human rac1 genomic
	13	116.4	33.4	3740	22	AAH07296	Human reproductive
	14	113	32.5	1017	22	AAH22397	Human rac1 gene re
C	15	95.4	27.4	201	25	AAK39132	Bovine EST associa
	16	63.8	18.3	68	16	AAH22394	Human gene signatu
C	17	60.8	17.5	64	25	ABE78455	Tumour suppressio
	18	60.8	17.5	64	25	ABE79002	Human oligonucleot
	19	52.4	15.1	162	24	ABK45998	CDNA encoding colo
C	20	52.4	15.1	605	21	AAH16478	Human colon cancer
	21	52.4	15.1	1022	22	AAH22398	Human rac1 gene re
C	22	50.8	14.6	605	24	ABJ38065	Human colon tumour
	23	50	14.4	50	24	ABJ01605	Human leukocyte ge
C	24	43.2	12.4	424	25	ABK46053	Bovine EST associa
	25	43	12.4	18660	22	AAK79108	Human immune/haema
	26	43	12.4	38844	24	AAH40285	Genomic DNA encodi
	27	42.2	12.1	640681	24	AAH292787	Buchnera sp. genom
C	28	42	12.1	15261	22	AAH07498	Human reproductive
	29	41.8	12.0	9817	24	ABJ33368	Human immune syste
	30	41.4	11.9	6464	24	ABJ32514	Human immune syste
C	31	40.8	11.7	5852	12	AAQ11710	Dictyostellium plas
	32	40.4	11.6	501	20	AAK97621	Extended human sec
C	33	40.2	11.6	185371	24	ABT10718	Human breast cancer
	34	40	11.5	534	23	AAK54457	Human prostate exp
C	35	39.6	11.4	1494	22	AAK86444	Human immune/haema
	36	39.6	11.4	2101	22	AAK86445	Human immune/haema
C	37	39.6	11.4	17918	24	AAH61419	Human gene regulat
C	38	39.4	11.3	6059	24	ABJ32813	Human immune syste
	39	39.4	11.3	6409	22	AAH46495	Tumour suppressor
C	40	39.2	11.3	277	25	ABX47508	Bovine EST associa
	41	39.2	11.3	375	25	AAK49849	Bovine EST associa
C	42	39.2	11.3	859	17	ABT32572	Microspore-specific
C	43	39.2	11.3	2039	17	AAH32569	Rapessed microspor
	44	39.2	11.3	6621	24	ABJ32913	Human immune syste
C	45	39.2	11.3	18671	22	AAK90763	Human digestive sy

ALIGNMENTS

RESULT 1	
AAH22399	
ID	AAH22399 standard; DNA; 348 Bp.
XX	
XX	AAH22399;
XX	
XX	
XX	
DT	22-AUG-2001 (first entry)
XX	
DE	Human rac1 gene related nucleotide sequence #3.
XX	
DE	Identification; toxic; hepatotoxic; differential gene expression;
KW	NSAID; non-steroidal antiinflammatory drug; ds.
KW	
XX	
OS	Homo sapiens.
XX	
PN	WO200138579-A2.
XX	
PD	31-MAY-2001.
XX	
XX	
PF	21-NOV-2000; 2000WO-US32049.
XX	
XX	
PPR	22-NOV-1999; 99US-0166923.
PPR	18-FEB-2000; 2000US-0183531.
PPR	20-NOV-2000; 2000US-0717321.
XX	
XX	
PA	(CURA-) CURAGEN CORP.
XX	
XX	
PFI	Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX	
DR	WPI; 2001-355948/37.
XX	
XX	

12

XX DE Lung cancer associated polynucleotide sequence SEQ ID 255.
XX DE Human; lung cancer associated protein; neuroprotective; cytostatic;
XX KW cardioactive; immunomodulatory; muscular active; vulnerary;
XX KW gastrointestinal; nephrotropic; antiinfective; gynecological;
XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX KW proliferative disorder; wound healing; infectious disease; ds.
XX OS Homo sapiens.
XX PN WO200055190-A2.
XX PD 21-SEP-2000.
XX XX 08-MAR-2000; 2000WO-US05919.
XX PF 12-MAR-1999; 99US-0124270.
XX XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX XX Ruben SM;
XX PI
XX XX WPI; 2000-587514/55.
XX DR P-PSDB; AAB58360.
XX XX
XX PT Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer -
XX PS Claim 1; Page 716-717; 1425pp; English.
XX XX
XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX CC associated proteins and polynucleotide sequences, their agonists, and
XX CC antagonists may have neuroprotective; cytostatic; cardioactive;
XX CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX CC general; nephrotropic; antiinfective; gynecological; or antibacterial
XX CC activity. The invention also includes antibodies specific for the
XX CC protein or polynucleotide sequences. The lung cancer associated
XX CC polynucleotide sequences may be used for detection of lung cancer,
XX CC chromosome identification, as chromosome markers, and for numerous other
XX CC diagnostic or research purposes. The proteins may be used to treat
XX CC disorders such as neural, immune, muscular, reproductive,
XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX CC disorders. The proteins may also be used in the treatment of wounds and
XX CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
XX CC peptide AAB58549 are used in the course of the invention for the
XX CC identification and characterisation of the polynucleotide and protein
XX CC sequences.
XX PS Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
XX SQ
Query Match 60.2%; Score 209.4; DB 21; Length 2051;
Best Local Similarity 82.6%; Pred. No. 1.4e-42;
Matches 285; Conservative 0; Mismatches 36; Indels 24; Gaps 3;
4 AACATCACTGCTTGGCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCTAGA 63
1666 AACATCACTGCTTGGCAGAGTACCGACACTGTCACCTTGACCAATCTGA-CCCTCTTTA 1724
64 CCTCACCCAGCGGACACATGCTCCGGTA-----CCTTGGGT 102
1725 CCTCGCCACGCGGACACAGCCCTCTGTAGTCGCTTTGTGCTATGTGATGTTCCCTTGGGT 1784
103 CTGTGAGGTTCTGTCAA--CGCGTAGTGTAAAGCGCGTCTGTACACCTAACTCACTGG 160
1785 CTGTGAGGTTCTGTAAACTGTGTAGTGTGACGATGTTCTGTACAACTTAACTCACTGG 1844
161 CAGAAACACAGTGTGGGCCCTTTCGACCACTAGACAAACTTTTTCAAATTGACAGTTGC 220
1845 CGAAGATACAGCTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATTGACAGTTGC 1904

QY 221 AGAATTGTGGAGTGTTTTACATTGAICTTTTGTCTAATGCAGTTAGCAGTATGTTTGA 280
DB 1905 AGAATTGTGGAGTGTTTTACATTGATCTTTTGTCTAATGCAGTTAGTATGTTTGA 1964
QY 281 TGTATGACTTAAATATCCCTTGAATCATAAAAAAATAAAAAA 325
DB 1965 TGTATGACTTAAATATCCCTTGAATCATAAAAAAATAAAAAA 2009

RESULT 4

ABT17030
ID ABT17030 standard; DNA; 28567 BP.
XX AC ABT17030;
XX DT 03-APR-2003 (first entry)
XX DE Human MP21 gene Racl SEQ ID No 4.
XX KW Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
XX KW cell proliferation disorder; MP21; gene; ds.
XX OS Homo sapiens.
XX PN WO2003006990-A1.
XX PD 23-JAN-2003.
XX PF 10-JUL-2002; 2002WO-US21549.
XX PR 12-JUL-2001; 2001US-305017P.
XX PR 10-OCT-2001; 2001US-328491P.
XX PR 15-FEB-2002; 2002US-357452P.
XX PA (EXEL-) EXELIXIS INC.
XX PI Friedman L, Plowman GD, Belvin M, Li D, Funke RP;
XX DR WPI; 2003-221779/21.
XX DR P-PSDB; ABJ19756.
XX PT Identifying candidate p21 pathway modulator, by contacting an assay
XX PT system having modifiers of p21 polypeptide or gene with a test agent to
XX PT provide a reference activity in system and detecting test agent-biased
XX PT activity -
XX PS Examples; Page 56-72; 199pp; English.
XX CC The invention relates to a novel method for identifying a candidate p21
XX CC pathway modulating agent. The novel method comprises contacting an assay
XX CC system, comprising a purified MP21 polypeptide (modifier of p21) or
XX CC nucleic acid, with a test agent under conditions, so that but for the
XX CC presence of a test agent, the assay system provides a reference activity
XX CC and detection of test agent-biased activity of the assay system. The
XX CC novel method of the invention is useful for identifying a candidate p21
XX CC pathway modulating agent. The invention also includes a method for
XX CC modulating the p21 pathway of a cell, and a method for diagnosing a
XX CC disease e.g. cancer in a patient. The identified modulators are useful in
XX CC diagnosis, therapy and pharmaceutical development. The modulators are
XX CC useful in a variety of diagnostic and therapeutic applications including
XX CC angiogenic, apoptotic and cell proliferation disorders. This
XX CC polynucleotide sequence represents a gene encoding an MP21 protein of the
XX CC invention.
XX SQ Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other;
Query Match 58.5%; Score 203.6; DB 25; Length 28567;
Best Local Similarity 78.8%; Pred. No. 7.5e-41;
Matches 290; Conservative 0; Mismatches 54; Indels 24; Gaps 3;
4 AACATCACTGCTTGGCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCTAGA 63

Db 27782 AACATCACTGCTTCCGAGATTACCGACACTGCTCACTTACCAATACTGA-CCCTCTTTA 27840
 QY 64 COTACCCACGCGGACACATGTTCCGGTA-----CCTTGGGT 102
 Db 27841 COTGCCACGCGGACACAGCCTCCTGTAGTGGCTTTCCTATTGATGTTCTTGGGT 27900
 QY 103 CTGTGAGTCTGTCAA--GCCTAGTGTAAAGCGCTTCTGTACAACTTAACCTCACTGG 160
 Db 27901 CTGTGAGTCTGTAACTGTCTAGTGTGACGATGTTCTGTACAACTTAACCTCACTGG 27960
 QY 161 CAAGAACACAGTGTGGGCTTTCGACCACTAGAACAAACTTTTCAATTGACAGTTGC 220
 Db 27961 CGAGATACAGCTGGGACCTTCAGCACTACACAGATTTTAAATTGACAGTTGC 28020
 QY 221 AGAATTGTGAGTGTGTTTACATGATCTTTTGTCTAATGAGTGTAGCAGTATGTTTGA 280
 Db 28021 AGAATTGTGAGTGTGTTTACATGATCTTTTGTCTAATGAGTGTAGCAGTATGTTTGA 28080
 QY 281 TGTATGACTTAATAATCCTTGAATCATTAATAAAAAAAATGCTTTTGGACTTG 340
 Db 28081 TGTATGACTTAATAATCCTTGAATCATGACTGTGTAATACTGGTGTGTTTGGACTTG 28140
 QY 341 AAAAAAA 348
 Db 28141 ATGAACAA 28148

RESULT 5
 ID AAH22400 standard; DNA; 352 BP.
 XX AC AAH22400;
 XX AC AAH22400;
 DT 22-AUG-2001 (first entry)
 XX Human rac1 gene related nucleotide sequence #4.
 XX Identification; toxic; hepatotoxic; differential gene expression;
 KW NSAID; non-steroidal antiinflammatory drug; ds.
 XX Homo sapiens.
 XX WO200138579-A2.
 XX 31-MAY-2001.
 XX 21-NOV-2000; 2000WO-US32049.
 XX 22-NOV-1999; 99US-0166923.
 PR 18-FEB-2000; 2000US-0183531.
 PR 20-NOV-2000; 2000US-0717321.
 XX (CURA-) CURAGEN CORP.
 PA Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
 PI WPI; 2001-355948/37.
 XX Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels -
 XX Disclosure; Page 8-9; 76pp; English.
 XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to

CC the expression of the nucleic acid sequence in an reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent in known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.
 XX Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other;
 SQ
 Query Match 54.0%; Score 187.8; DB 22; Length 352;
 Best Local Similarity 77.6%; Pred. No. 2.1e-37;
 Matches 274; Conservative 0; Mismatches 57; Indels 22; Gaps 3;
 QY 17 TCCAGAGTACCAACACTGTCATGTGATGATGCGCCGCCCTCTAGACCTCACCCACGG 76
 Db 1 TCCAGAGTACCGACACTGTCATGACCAATACG-ACCCTCTTTACCTCGCCACGG 59
 QY 77 GACATGCTTCG-----GTACCTTTGGGTCTGTGAGTCTCTGTC 117
 Db 60 GACACCGCTCTGCTGCTGCTTGGCTATTGATGTTCCITTTGGTCTGTGAGTCTCTGTA 119
 QY 118 AA--GCGTGTAGTCTAACCGCGTCTGTACCACTTAACCTCACTGCGCAAGACACAGTGT 175
 Db 120 AACTGTGTAGTGTGACGATGTTCTGTACAACTTAACCTCACTGCGGAGAAATACAGGTG 179
 QY 176 GGGCTTTTCGACACTAGAACAACTTTTCAATTGACAGTGTGCAAGATTGGAGTGT 235
 Db 180 GGACCTTCAGCCACTACACAGAAATTTTAAATTGACAGTGTGCAAGATTGGAGTGT 239
 QY 236 TTTTACATTGATCTTTTGTCTAATGACAGTGTAGCAGTATGTTTGCATGTATGACTTAATA 295
 Db 240 TTTTACATTGATCTTTTGTCTAATGCAATTAGCATTATGTTTGCATGTATGACTTAATA 299
 QY 296 ATCTTTGAATCATTAATAAAAAAAATGCTTTTGGAACTTGAATAAAAA 348
 Db 300 ATCTTTGAATCATACGACTGGTAAATGCTGTTTGTGAGACTTGTATGAACAA 352
 RESULT 6
 ID ABK83623 standard; cDNA; 1232 BP.
 XX AC ABK83623;
 XX AC ABK83623;
 DT 14-AUG-2002 (first entry)
 XX Human cDNA differentially expressed in granulocytic cells #194.
 XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX Homo sapiens.
 OS WO200228999-A2.
 XX WO200228999-A2.
 XX 11-APR-2002.
 XX 03-OCT-2001; 2001WO-US30821.
 XX 03-OCT-2000; 2000US-237189P.
 PR (GENE-) GENE LOGIC INC.
 PA Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 PI WPI; 2002-435328/46.
 XX DR

XX Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 194; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulation of (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC or allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
Query Match 53.9%; Score 187.4; DB 24; Length 1232;
Best Local Similarity 80.9%; Pred. No. 3.6e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
QY 4 AACATCACTGTCTTGGCCAGAGTACCAACACTGTCATGTGATTGATCGCGCCCTCTAGA 63
DB 905 AACATCACTGTCTTGGCCAGATACCGACACTGTCACTTGACCAATACTGA-CCCTCTTTA 963
QY 64 CCTCACCCACGGGACACATGCTTCGGTA-----CCTTTGGGT 102
DB 964 CCTCGCCACGGGACACACAGCCTCTCTGTAGTCGCTTTCCTATTGATGTTCTTTGGGT 1023
QY 103 CTGTGAGGTTCTGTCAA--GCGCTAGTGTACGCGCTTCTGTACAACTTAACCTACTGG 160
DB 1024 CTGTGAGGTTCTGTAAATGTGCTAGTGTGACGATGTTCTGTACAACTTAACCTACTGG 1083
QY 161 CAAGAACACAGTGTCTGGCCCTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGC 220
DB 1084 CGAAGATACAGCGTGGAGCCCTTCACCCACTACACAGAAATTTTAAATTTGACAGTTGC 1143
QY 221 AGAATTTGTGAGGTTTTCATTGATGATCTTTCTTAATGACAGTTAGCATGTTTGTGA 280
DB 1144 AGAATTTGTGAGGTTTTCATTGATGATCTTTTCTAATGCAATTAGCATTTATTTGTGA 1203
QY 281 TGTATGACTTAATAATCTTGAAT 305
|||||

DB 1204 TGTATGACTTAATAAATCCTTGGAW 1228
RESULT 7
ABN95134
ID ABN95134 standard; DNA; 1232 BP.
XX
AC ABN95134;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #1632 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX
XX Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
XX liver tissue sample -
PS Claim 1; SEQ ID NO 1632; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
Query Match 53.9%; Score 187.4; DB 24; Length 1232;
Best Local Similarity 80.9%; Pred. No. 3.6e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
QY 4 AACATCACTGTCTTGGCCAGAGTACCAACACTGTCATGTGATTGATCGCGCCCTCTAGA 63
DB 905 AACATCACTGTCTTGGCCAGATACCGACACTGTCACTTGACCAATACTGA-CCCTCTTTA 963
QY 64 CCTCACCCACGGGACACATGCTTCGGTA-----CCTTTGGGT 102
DB 964 CCTCGCCACGGGACACACAGCCTCTCTGTAGTCGCTTTCCTATTGATGTTCTTTGGGT 1023
QY 103 CTGTGAGGTTCTGTCAA--GCGCTAGTGTACGCGCTTCTGTACAACTTAACCTACTGG 160
DB 1024 CTGTGAGGTTCTGTAAATGTGCTAGTGTGACGATGTTCTGTACAACTTAACCTACTGG 1083

QY 161 CAAGACACAGTGTGGCCCTTTCACACCTAGAACAACTTTTCAATTGACAGTTGC 220
 Db 1084 CGAGAAATACAGCGTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATTGACAGTTGC 1143
 QY 221 AGAATTGGAGTGTGTTTACATTCATCTTTTGTCTAATGACAGTGTAGCAGTATGTTTGA 280
 Db 1144 AGAATTGGAGTGTGTTTACATTCATCTTTTGTCTAATGACAGTGTAGCAGTATGTTTGA 1203
 QY 281 TGTATGACTTAATAAATCCTTGAAT 305
 Db 1204 TGTATGACTTAATAAATCCTTGAAT 1228

RESULT 8
 ID AAS65569 standard; cDNA; 1318 BP.
 XX
 AC AAS65569;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #1373.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 EN WO200175067-A2.
 XX
 PD 11-OCT-2001.

XX 30-MAR-2001; 2001WC-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PI Dmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG01382.
 DR
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 1; SEQ ID No 1373; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
 Query Match 53.9%; Score 187.4; DB 23; Length 1318;
 Best Local Similarity 80.9%; Pred. No. 3.7e-37;
 Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
 QY 4 AACATCACTGCTTCCGACAGTACCAACACTGTCTATGTGATGTGCGCCGCCCTCTAGA 63
 Db 991 AACATCACTGCTTCCGACAGTACCAACACTGTCTATGTGATGTGCGCCGCCCTCTAGA 1049
 QY 64 CCTCACCCACGCGGACACATGCTTCGGTA-----CCTTTGGGT 102
 Db 1050 CCTCGCCACGCGGACACATGCTTCGGTA-----CCTTTGGGT 1109
 QY 103 CTGTGAGGTTCTGTCAA--GCGCTAGTGTCTAACGCGGTTCTGTCAACACCTAACTCACTGG 160
 Db 1110 CTGTGAGGTTCTGTAACTAGTGTCTAGTGTCTGTACAGATGTTCTGTACAACTTAACACTG 1169
 QY 161 CAAGAACACAGTGTGGGCTTTCGACCACTAGAACAACTTTTCAATTGACAGTTGC 220
 Db 1170 CGAGAAATACAGCGTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATTGACAGTTGC 1229
 QY 221 AGAATTGGAGTGTGTTTACATTCATCTTTTGTCTAATGACAGTGTAGCAGTATGTTTGA 280
 Db 1230 AGAATTGGAGTGTGTTTACATTCATCTTTTGTCTAATGACAGTGTAGCAGTATGTTTGA 1289
 QY 281 TGTATGACTTAATAAATCCTTGAAT 305
 Db 1290 TGTATGACTTAATAAATCCTTGAAT 1314

RESULT 9
 ABX46267/C
 ID ABX46267 standard; cDNA; 353 BP.
 XX
 AC ABX46267;
 XX
 DT 21-FEB-2003 (first entry)
 XX
 DE Bovine EST associated with lactation/muscle/fat deposition #11432.
 XX
 DE Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 XX US2002137139-A1.
 PN
 XX 26-SEP-2002.
 PD
 XX 24-SEP-2001; 2001US-0960352.
 PF
 XX 12-JAN-1999; 99US-115707P.
 PR 11-JAN-2000; 2000US-0480902.
 PR
 XX (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;
 PI
 XX WPI; 2003-110599/10.
 XX
 XX New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and
 PT analysis, cattle breeding, or for genetically improving cattle
 XX
 XX Claim 2; SEQ ID No 11432; 245pp; English.
 PS
 XX The invention relates to a purified nucleic acid molecule associated with
 CC

XX	08-JAN-2001	(first entry)
DT		
XX		
DE		
XX		
XX	Mouse Rab2 nucleotide sequence #1.	
KW		
KW	Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;	
KW	antiallergic; antisthmatic; nootropic; neuroprotective; anticonvulsant;	
KW	vulnerable; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;	
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;	
KW	digestion disorder; wound healing disorder; gene therapy; ss.	
XX		
OS	Mus sp.	
XX		
XX	WO200043419-A2.	
XX		
PD	27-JUL-2000.	
XX		
XX	20-JAN-2000; 2000WO-US01431.	
PF		
XX	20-JAN-1999; 99US-0116534.	
PR	26-JAN-1999; 99US-0117274.	
PR	26-JAN-1999; 99US-0117308.	
PR	26-JAN-1999; 99US-0117309.	
PR	26-JAN-1999; 99US-0117312.	
PR	26-JAN-1999; 99US-0118177.	
PR	01-FEB-1999; 99US-0118178.	
PR	01-FEB-1999; 99US-0118179.	
PR	09-FEB-1999; 99US-0119286.	
PR	11-FEB-1999; 99US-0119998.	
PR	11-FEB-1999; 99US-0119759.	
XX		
XX	(RIGE-) RIGEL PHARM INC.	
PA		
XX		
PI	Luo Y;	
XX		
XX	WPI; 2000-482908/42.	
DR		
XX		
PT	New nucleic acids encoding Exo proteins which are useful in the	
PT	diagnosis, treatment or prevention of exocytosis-mediated disorders	
PT	such as asthma, inflammation and allergies -	
XX		
XX	Disclosure; Page 271-272; 305pp; English.	
PS		
XX		
CC	The present sequence encodes a polypeptide which is associated with	
CC	the exocytosis pathway. cDNA molecules encoding proteins involved in	
CC	exocytosis have been isolated by yeast one-hybrid and two-hybrid	
CC	screening. Novel proteins, termed Exo proteins, have been identified that	
CC	interact with known exocytosis-associated proteins such as GS27, alpha	
CC	snap, unc18-1, vamps, snap-23, and the rab family of proteins.	
CC	Exo proteins and their agonists and antagonists are useful in the	
CC	diagnosis, treatment or prevention of exocytosis-mediated disorders	
CC	such as asthma, inflammation, allergies, Chediak-Higashi Syndrome	
CC	(CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,	
CC	diabetes, digestion disorders and wound healing disorders.	
CC	The nucleic acids, antagonists or agonists of Exo proteins are useful	
CC	in gene therapy. The nucleic acids are also useful for generating	
CC	transgenic or knock-out animals which can be used in the	
CC	development and screening of therapeutically useful reagents.	
XX		
XX	Sequence 422 BP; 95 A; 92 C; 83 G; 148 T; 4 other;	
QY		
QY	Query Match 41.7%; Score 145.2; DB 21; Length 422;	
DB	Best Local Similarity 82.4%; Pred. No. 9.1e-27;	
DB	Matches 211; Conservative 0; Mismatches 20; Indels 25; Gaps 3	
QY	21 AGAGTACCAACACTGTCATGTGATTCGCGGCCCTCTAGACCTCACCCACGCGACA 80	
DB	169 ACAGTACCAACACTGTCATGTGACTAATGCTG--CCTCTANACCTCGCCACGTGACA 225	
QY	81 CATGCTCCGGTA-----CCTTTGGGTCGTGAGGTCCTGTCAA 119	
DB	226 GACGCTTCCTGTCATGGCTCTGCCTANAGATGTTCTCTGGGTCGTGAGGTCCTGTCAA 285	

QY 90 GTACTCTTGGCTGTGAGGTTCTGTCAAGGCTAGTGTAAAGCGCTTCTGTACAACC 149
 Db |||||
 123 GGTACTCTTGGCTGTGAGGTTCTGTCAAGGCTAGTGTAAAGCGCTTCTGTACAACC 64
 QY 150 TAACTCAGTGGCAAGACACAGTGTGGGCTTTGACACACTAGAACAACTTTTTCAA 209
 Db |||||
 63 TAACTCAGTGGCAAGACACAGTGTGGGCTTTGACACACTAGAACAACTTTTTCAA 4
 QY 210 TTG 212
 Db ||||
 3 TTG 1
 RESULT 13
 AAL07296
 ID AAL07296 standard; DNA; 3740 BP.
 XX
 AC AAL07296;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen DNA SEQ ID NO: 9984.
 XX
 KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0232403.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.

```

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 984; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
XX Sequence 3740 BP; 1188 A; 752 C; 687 G; 1113 T; 0 other;
SQ
Query Match 33.4%; Score 116.4; DB 22; Length 3740;
Best Local Similarity 82.0%; Pred. No. 2.4e-19;
Matches 146; Conservative 0; Mismatches 31; Indels 1; Gaps 1;
QY 147 ACCTAACTCACTGGCAGAACACAGTGTGGGCTTTTGGACCACTAGAACAAACTTTTTT 206
Db 2667 ACCTTAAGCACTGGCAGAAATATATACACTGGGCTTTTCACTAGAACAAA-TATTTT 2725
QY 207 CAATTGACAGTTGCAGAAATGTGGAGTGTGTTTACATTTGATGCTTAATGAGTTAG 266
Db 2726 AATGACAGATTCAGAAATGTGGGTATTTTACATTTGATGCTTAATGCAATCAG 2785
QY 267 CAGTATGTTTGCATGTATGACCTTAATAATCCTTGAATCAATAAAAAAAAAAAAA 324
Db 2786 CAATGTGTTTGCACATGGAATTAATAATCCTTGAATCAATAATAATAATAATAA 2843
RESULT 14
AAH22397
ID AAH22397 standard; DNA; 1017 BP.
XX
XX AAH22397;
XX
XX 22-AUG-2001 (first entry)
XX
XX Human rac1 gene related nucleotide sequence #1.
XX
XX Identification; toxic; hepatotoxic; differential gene expression;
XX NfAID; non-steroidal antiinflammatory drug; ds.

```

```

XX OS Homo sapiens.
XX PN WO200138579-A2.
XX PD 31-MAY-2001.
XX PF 21-NOV-2000; 2000WO-US32049.
XX PR 22-NOV-1999; 9US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX
XX WPI; 2001-355948/37.
XX
XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
XX Disclosure; Page 7-8; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in a reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
XX Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;
SQ
Query Match 32.5%; Score 113; DB 22; Length 1017;
Best Local Similarity 100.0%; Pred. No. 1.2e-18;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGAACATCACTGTCTTGGCCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 60
Db 905 TGAACATCACTGTCTTGGCCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 964
QY 61 AGACCTCACCCAGCGGACACATGCTTCGGTACCTTGGTCTGTGAGTTC 113
Db 965 AGACCTCACCCAGCGGACACATGCTTCGGTACCTTGGTCTGTGAGTTC 1017
RESULT 15
ABX39192/c
ID ABX39192 standard; cDNA; 201 BP.
XX
XX ABX39192;
XX
XX 20-FEB-2003 (first entry)
XX
XX Bovine EST associated with lactation/muscle/fat deposition #4357.
XX
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
XX muscle deposition; fat deposition; genome mapping; gene identification;
XX gene analysis; cattle breeding.
XX
XX Bos Taurus.
XX

```


PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
XX 24-SEP-2001; 2001US-0960352.
PF
XX 12-JAN-1999; 99US-115707P.
PR 11-JAN-2000; 2000US-0480902.
XX
XX (BYATT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
XX WPI; 2003-110599/10.
XX
XX New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle
XX
PS Claim 2; SEQ ID No 4357; 245pp; English.
XX
XX The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docID=20020137139.
XX
SQ Sequence 201 BP; 67 A; 34 C; 23 G; 77 T; 0 other;

Query Match 27.4%; Score 95.4; DB 25; Length 201;
Best Local Similarity 76.5%; Pred. No. 1.9e-14;
Matches 117; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 174 TTGGGCTTTTCGACCACTAGACAACTTTTTCATTTGACAGTTGAGATTTGTGGAGT 233
Dd |||||
193 TTGAACCTTTTCAAAAATTAGAACATAATTTTAAATTGACAGTTGAGATTTGTGGAGT 134
QY 234 GTTTTACATTTCTTCTTAATGAGTATGAGTATCTTTTGCATGTATGACTTAAT 293
Dd |||||
133 GATTTTAGATTTGATTTTTCATACGTGATTAAGATGATCTTTTGAAGGTATCCCTTAAT 74
QY 294 AAATCCTTGAATCATAAAAAATAAAAAAT 326
Dd |||||
73 AAATTTTGACCTTCAGTATAAAGATAAAAT 41

Search completed: November 23, 2003, 11:09:41
Job time : 101.194 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:18:57 : Search time 1304.05 Seconds
(without alignments)
10917.162 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348

Sequence: 1 tgaacatcactgctgcc.....cttgggaacttgaaaaaaa 348

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pla.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	348	100.0	348	6	AX163753	AX163753 Sequence
C	348	100.0	1266	6	AX163738	AX163738 Sequence
3	320	92.0	217700	2	AC106124	AC106124 Rattus no
4	230.2	66.1	2281	10	BC051053	BC051053 Mus muscu
5	228.2	65.6	2319	10	BC003828	BC003828 Mus muscu
6	221	63.5	269081	2	AC068493	AC068493 Mus muscu
C	216.8	62.3	192498	2	AC105979	AC105979 Mus muscu
C	216.8	62.3	230015	2	AC132602	AC132602 Mus muscu
9	207.4	59.6	2302	9	BC050687	BC050687 Homo sapi
10	203.6	58.5	28567	9	HSAL132695	HSAL132695 Homo sapi
11	203.6	58.5	212827	9	AC009412	AC009412 Homo sapi
12	192.4	55.3	2315	9	AK054993	AK054993 Homo sapi
13	187.8	54.0	352	6	AX163754	AX163754 Sequence
14	187.4	53.9	1232	6	AX408985	AX408985 Sequence
15	187.4	53.9	1232	9	HUMPO2ST9	D25274 Homo sapien
C	176.4	50.7	87616	2	AC139405	AC139405 Homo sapi
17	176.4	50.7	137625	9	AC104663	AC104663 Homo sapi
C	175.6	50.5	240973	2	AC123247	AC123247 Rattus no
19	175.6	50.5	241048	2	AC129824	AC129824 Rattus no
C	174.8	50.2	5544	9	AF542527	AF542527 Homo sapi
C	170.6	49.0	262	11	G31709	G31709 SWS2233 Er
C	154	44.3	455	11	G26995	G26995 human STS S
C	123	35.3	123	6	AX163737	AX163737 Sequence
24	121.4	34.9	228121	2	AC133022	AC133022 Rattus no
C	121.4	34.9	239768	2	AC112582	AC112582 Rattus no
26	121.4	34.9	245468	2	AC130746	AC130746 Rattus no
C	116.4	33.4	110816	9	AC002404	AC002404 Human Chr
28	113	32.5	1017	6	AX163751	AX163751 Sequence
C	110.6	31.8	101584	9	CNS01DS5	AL121655 BAC sequ
C	110.6	31.8	155943	9	AC012364	AC012364 Homo sapi
31	84.6	24.3	174316	2	AC022647	AC022647 Homo sapi
32	84.6	24.3	217249	9	AC009902	AC009902 Homo sapi
33	81.2	23.3	118648	9	AL138742	AL138742 Human DNA
34	81.2	23.3	131095	2	AL590071	AL590071 Homo sapi
C	79.2	22.8	170839	2	AC133467	AC133467 Mus muscu
C	75.2	21.6	237985	2	AC125754	AC125754 Rattus no
37	75.2	21.6	247478	2	AC097964	AC097964 Rattus no
38	72	20.7	156879	10	AL626786	AL626786 Mouse DNA
C	67	19.3	340701	2	AC120633	AC120633 Rattus no
40	66.8	19.2	62656	2	AC100110	AC100110 Mus muscu
C	66.8	19.2	178482	2	AC101810	AC101810 Mus muscu
42	60.8	17.5	64	6	AX522492	AX522492 Sequence
43	53.4	15.3	100697	5	AL672072	AL672072 Zebrafish
44	52.4	15.1	162	6	AX397334	AX397334 Sequence
45	52.4	15.1	1022	6	AX163752	AX163752 Sequence

ALIGNMENTS

RESULT 1
AX163753
LOCUS AX163753 348 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 17 from Patent WO0138579.
ACCESSION AX163753
VERSION AX163753.1 GI:14544859
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Gould-Rothberg, B.E., Dipippo, V.A., Ramse, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential

Pred. No. is the number of results predicted by chance to have a

QY	61	AGACCTCACCACGGGACACATGTTCCGGTACCTTTGGGTCTGTGAGGTTCCTGTCAAG	120
Db	297	AGACCTCACCACGGGACACATGTTCCGGTACCTTTGGGTCTGTGAGGTTCCTGTCAAG	238
QY	121	CGCTAGTGCTAACGCCGTTCTGTAAACCTTAACCTCACTGGCAAGAACACAGTGTGGGCC	180
Db	237	CGCTAGTGCTAACGCCGTTCTGTAAACCTTAACCTCACTGGCAAGAACACAGTGTGGGCC	178
QY	181	TTTCGACCACTAGACAACAATTCTTCAATTCACAGTTGCAGAAATCTGCAGATGTTTTTA	240
Db	177	TTTCGACCACTAGACAACAATTCTTCAATTCACAGTTGCAGAAATCTGCAGATGTTTTTA	118
QY	241	CATTGATCTTTTGCCTAATGCAGTAGCAGTAGTGTTCATGTATGACTTAATAAATCCT	300
Db	117	CATTGATCTTTTGCCTAATGCAGTAGCAGTAGTGTTCATGTATGACTTAATAAATCCT	58
QY	301	TGAATCATNAAAAAAAATGCTTTTGAACCTTGAIAAAAAA 348	
Db	57	TGAATCATNAAAAAAAATGCTTTTGAACCTTGAIAAAAAA 10	
RESULT 3		217700 bp DNA linear HTG 10-MAY-2003	
AC106124		Rattus norvegicus clone CH230-119E11, *** SEQUENCING IN PROGRESS	
LOCUS		***, 11 unordered pieces.	
DEFINITION		AC106124	
ACCESSION		AC106124_6 GI:30521557	
VERSION		HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.	
KEYWORDS		Rattus norvegicus (Norway rat)	
SOURCE		Rattus norvegicus	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE		1 (bases 1 to 217700) Muzny,D.,Marie,, Metckr,M.Lee,, Abramzon,S., Adams,C., Alder,J.,, Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebech,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Bulay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Clelland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,K., Garcia,A., Garner,T., Garza,M., Georgievski,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jovivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowis,C., Kraft,C.I., Lebow,H., Levay,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A., Manungu,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,B., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwokeleneh,O., Okwuonu,G., Olarnpunagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C., Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.-I., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Rem,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojlas,A., Rose,M., Rose,R., Ruiz,S.J.,,	

Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Staimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villaesana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, K., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

On May 10, 2003 this sequence version replaced gi:25094619. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHL
Center clone name: CH230-119B11

```

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185725 bases at least Q40
Consensus quality: 190140 bases at least Q30
Consensus quality: 192842 bases at least Q20
Estimated insert size: 194758; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

```

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 11 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

Accession	Contig	Length (bp)
1	109797	contig of 109797 bp in length
109798	109897	gap of unknown length
109898	166934	contig of 57037 bp in length
166935	167034	gap of unknown length
167035	176384	contig of 9350 bp in length
176385	176484	gap of unknown length
176485	206327	contig of 23843 bp in length
206328	206427	gap of unknown length

[illegible]

RECIT. T. A.

RESUL 4
BC051053

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS
SOURCE

SOURCE
ORGANIS

ORGANIS

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REPORT	COMMENT
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

1100

Db	2247	TTAATTAATCCTTGATCATATAAAAAAAAAAAAAA	2281
RESULT 5			
BC003828			
LOCUS	BC003828	2319 bp	linear
DEFINITION	Mus musculus RAS-related C3 botulinum substrate 1, mRNA (cDNA clone MGC:6235 IMAGE:359357), complete cds.		
ACCESSION	BC003828		
VERSION	BC003828.1	GI:13277917	
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

REFERENCE

1^a (bases 1 to 2319)
Srausberg R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L.H., Shennen C.F., Bhat, N.K., Altschul, S.F., Zeeberg, B., Bueter, K.H., Schaefer, C.P., Hsieh, F., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Drachtenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scaplehorn, M., Soares, N.B., Bonaldo, M.P., Casaventi, T.I., Thompson, J.C.

AUTHORS

TITLE
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL OF CELLULAR PHYSIOLOGY 190:233-241 (2002)
 TITLE: Human and mouse cDNA sequences
 AUTHORS: Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
 REFERENCE: 23388257
 MEDLINE: 23388257
 PUBMED: 12477932
 REFERENCE: 2 (bases 1 to 2319)
 AUTHORS: Strausberg, R.
 TITLE: Direct Submission
 JOURNAL: Submitted (28-FEB-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

```

COMMENT
Contact: MGC help desk
Email: cgap@remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: irak Plate: 11 Row: e Column: 22.
Location/Qualifiers
1. .2319
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:6235 IMAGE:3593957"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"

FEATURES
source

```

```
/lab host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. 2319
/genes="Rac1"
/notes="synonym: D5Erd559e"
/db_xref="LocusID:19353"
/db_xref="MGI:97845"
198..776
/codon_start=1
/product="Rac1 protein"
/protein_id="AAH03828.1"
/db_xref="GI:13277918"
/db_xref="LocusID:19353"
MVDKPNLGLWDYTAGQEDYDRPLSPQIDVFLICSLVSPASFNVRKWPYEV
HHCNTPLILGTLKLDKDDKOTIEKKRLKLTITYPQGLAMAKEIGAVKYLRCAL
TORGLKTVFDEAIVLVCPPVKKRKKCLLL"
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN
Query Match 65.6%; Score 228.2; DB 10; Length 2319;
Best Local Similarity 83.7%; Pred. No. 8.5e-48;
Matches 292; Conservative 0; Mismatches 33; Indels 24; Gaps 2;
QY 21 AGAGTACCAACACTGTCTATGATGATGCGCCCTCTAGACTCTACCCACGCGGACA 80
Db 1964 ACAGTACCAACACTGTCTATGATGATGCGCCCTCTAGACTCTACCCACGCGGACA 2020
QY 81 CATGCTCCGGTA-----CCTTGGGCTGTGAGGTTCTGTCAA 119
Db 2021 GACGCTTCCTGTCTATGCTGTGCTGTGAGGTTCTGTGAGGTTCTGTCAA 2080
QY 120 GCGCTAGTGTAAAGCGGTTCTGTACACCTAAGTCTGCTGACAGACACAGTGTGGG 179
Db 2081 CTGCTAGTGTAAAGCGGTTCTGTACACCTAAGTCTGCTGACAGACACAGTGTGGG 2140
QY 180 CTTTCGACCACTAGACAAACTTTTCAATGACAGTGTGAGGTTCTGTCAA 239
Db 2141 CTTTCGACCACTAGACAAACTTTTCAATGACAGTGTGAGGTTCTGTCAA 2200
QY 240 ACATTGATCTTTGCTAATGACAGTGTGAGGTTCTGTCAA 299
Db 2201 ACATTGATCTTTGCTAATGACAGTGTGAGGTTCTGTCAA 2260
QY 300 TTGAATCATATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 348
Db 2261 TTGAATCATATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2309
AC068493 269081 bp DNA linear HTG 15-MAY-2002
Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
AC068493
AC068493.10 GI:15148081
HTG: HTGS PHASE1; HTGS DRAFT.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269081)
Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A.,
Gordon M., Goltz J.S. and Kucherlapati R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 269081)
Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A.,
Gordon M., Goltz J.S. and Kucherlapati R.
Direct Submission
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
```

```
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPOC
Web site: http://www.hpcgg.org/Sequence/mouse.html
Contact: hpcg@mcg.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 268591 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 51730: contig of 51730 bp in length
* 51731 51750: gap of unknown length
* 51751 82423: contig of 30673 bp in length
* 82424 82443: gap of unknown length
* 82444 118158: contig of 35715 bp in length
* 118159 118178: gap of unknown length
* 118179 139441: contig of 21263 bp in length
* 139442 139461: gap of unknown length
* 139462 160638: contig of 21177 bp in length
* 160639 160658: gap of unknown length
* 160659 177611: contig of 16953 bp in length
* 177612 177631: gap of unknown length
* 177632 180991: contig of 13360 bp in length
* 180992 190111: gap of unknown length
* 190112 200129: contig of 9117 bp in length
* 200130 200148: gap of unknown length
* 200149 211882: contig of 11734 bp in length
* 211883 211902: gap of unknown length
* 211903 220773: contig of 8871 bp in length
* 220774 220794: gap of unknown length
* 220795 229643: contig of 8850 bp in length
* 229644 229663: gap of unknown length
* 229664 238339: contig of 8676 bp in length
* 238340 238359: gap of unknown length
* 238360 244848: contig of 6489 bp in length
* 244849 244868: gap of unknown length
* 244869 249725: contig of 4857 bp in length
* 249726 252523: contig of 2778 bp in length
* 252524 252543: gap of unknown length
* 252544 257520: contig of 4977 bp in length
* 257521 257540: gap of unknown length
* 257541 260396: contig of 2856 bp in length
* 260397 260416: gap of unknown length
* 260417 262709: contig of 2293 bp in length
* 262710 262729: gap of unknown length
* 262730 263934: contig of 1205 bp in length
* 263935 264196: gap of unknown length
* 264197 264216: contig of 242 bp in length
* 264217 264629: contig of 413 bp in length
* 264630 264649: gap of unknown length
* 264650 265608: contig of 959 bp in length
* 265609 265628: gap of unknown length
* 265629 266935: contig of 1307 bp in length
```

COMMENT

JOURNAL

Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0477C13

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 225083 bases at least Q40
Consensus quality: 225730 bases at least Q30
Consensus quality: 226267 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 228405; sum-of-contigs
Quality coverage: 13.11 in Q20 bases; agarose-fp
Quality coverage: 9.64 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 710: contig of 710 bp in length
711 810: gap of unknown length
811 1816: contig of 17506 bp in length
1817 18416: gap of unknown length
18417 57380: contig of 38964 bp in length
57381 57480: gap of unknown length
57481 118173: contig of 60693 bp in length
118174 118273: gap of unknown length
118274 201518: contig of 83245 bp in length
201519 201618: gap of unknown length
201619 202801: contig of 1183 bp in length
202802 204714: gap of unknown length
204715 204814: gap of unknown length
204815 208091: contig of 3277 bp in length
208092 208191: gap of unknown length
208192 213082: contig of 4891 bp in length
213083 213183: gap of unknown length
213183 230015: contig of 16833 bp in length.

FEATURES

source

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-477C13"

misc_feature

1..710

/note="assembly_name:Contig1"

811..1816

/note="assembly_name:Contig10"

18417..57380

/note="assembly_name:Contig11"

57481..118173

/note="assembly_name:Contig12"

118274..201518

/note="assembly_name:Contig13"

201619..202801

/note="assembly_name:Contig3"

202902..204714

/note="assembly_name:Contig6"

misc_feature

204815..208091

/note="assembly_name:Contig7"

misc_feature

208192..213082

/note="assembly_name:Contig8"

misc_feature

213183..230015

/note="assembly_name:Contig9"

BASE COUNT 63025 a 50551 c 50278 g 65213 t 948 others

ORIGIN

Query Match 62.3%; Score 216.8; DB 2; Length 230015;

Best Local Similarity 86.1%; Pred. No. 8.6e-45;

Matches 272; Conservative 0; Mismatches 22; Indels 22; Gaps 2;

29 AACACTGTCATGTGATGCGGCCCTCTAGACCTCACCACCGGACACATGCTTC 88

109281 AACACTGTCATGTGACTAATGCTG-CCCTCTAGACCTCCTCACCAGACAGCGTTC 109223

89 CGGTA-----CCTTGGGCTCTGAGGTTCTGCAAGCGCTAGT 127

109222 CTGTCGTGGCTCTGCTAGAGATGTTCTCTGGGGTCTGTGAGGTTCTGTCAACTGCTAGT 109163

128 GCTAACGGCGTCTGTGACAACTTAACCTCACTGCGAAGACACAGTGTTCGGGCTTCGAC 187

109162 GCTAATGCTGCTCTGTACAACTTAACCTCACTGCGAAGATACAGTGTTCGGGCTTCGAC 109103

188 CACTAGACAAACTTTTTTCAATTGACAGTTGACAGATTGTGGAGTGTTCACATGAT 247

109102 CACTAGACAAACTTTTTTCAATTGACAGTTGACAGATTGTGGAGTGTTCACATGAT 109043

248 CTTTGTCTAATGACAGTTAGCAGTATGTTTGTGATGATGACATTAATAATCCTTGAATCA 307

109042 CTTTGTCTAATGACAGTTAGCAGTATGTTTGTGATGATGACATTAATAATCCTTGAATCA 108983

308 TAAATAAATAAATAA 323

108982 TTAATAAATAAATAA 108967

RESULT 9

LOCUS BC050687 2302 bp mRNA linear PRI 11-APR-2003

DEFINITION Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho

IMAGE:6149377, mRNA, complete cds.

ACCESSION BC050687.1 GI:29792301

KEYWORDS MGC.

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 2302)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (08-APR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/PTP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

[illegible]

RESULT 11
AC009412 212827 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
DEFINITION AC009412
ACCESSION AC009412
VERSION AC009412.6 GI:14190769
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212827)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PubMed 9847074
REFERENCE 2 (bases 1 to 212827)
AUTHORS Hou, S., Maupin, R., Haakenson, W., Gregory, S. and Belter, R.
TITLE The sequence of Homo sapiens BAC clone RP11-425P5
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 212827)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On May 23, 2001 this sequence version replaced gi:13431187.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male

donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACE3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is CTD-2195P2, 200 bp overlap.
Actual start of this clone is at base position 1 of RP11-425P5;
actual end is at base position 212633 of RP11-425P5.

FEATURES	source
1. 212827	Location/Qualifiers
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="7"	
/map="7"	
/clone="RP11-425P5"	
/clone_lib="RPC1-11"	
1623..1696	
/rpt_family="MIR"	
1698..2006	
/rpt_family="Alu"	
2024..2337	
/rpt_family="Alu"	
3006..3093	
/rpt_family="L1"	
3453..3552	
/rpt_family="L2"	
4924..5046	
/rpt_family="ERV1"	
5318..5693	
/note="similar to EST BF996284 (NID:gi12402607)"	
5580..5845	
/rpt_family="Alu"	
5925..6277	
/rpt_family="MALR"	
6842..7041	
/note="similar to EST BF369507 (NID:gi1331532)"	
7364..7661	
/rpt_family="Alu"	
8029..8048	
/rpt_family="(CAAAA)n"	
8379..8678	
/rpt_family="Alu"	
8560..8762	
/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"	
8720..9031	
/rpt_family="Alu"	
9083..9560	
/note="similar to EST AA633603 (NID:g25556817) ae66b02.s1"	
9598..9899	
/rpt_family="Alu"	
9871..10098	
/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"	
10746..11044	
/rpt_family="L1"	
11075..11101	
/rpt_family="AT-rich"	
11102..11372	
/rpt_family="Alu"	
11411..11465	
/rpt_family="GA-rich"	
11552..11644	
/rpt_family="L2"	
11685..12193	
/rpt_family="L1"	
12194..12486	
/rpt_family="Alu"	
12487..13003	
/rpt_family="L1"	

[illegible]

```

QY 64 CCTCACCACGGGACACATGCTTCGGTA-----CCTTTGGGT 102
Db 2048 CCTGCCCCACGGGACACACGGCTCTGTAGTCGCTTTCCTTATGATGTTCTTTGGGT 2107
QY 103 CTGTGAGGTTCTGTCAA--GCCTAGTGCTAACGCCGTTCTGTACAACCTAACTCACTGG 160
Db 2108 CTGTGAGGTTCTGTAACTGTCTAGTGTGACGATGTTCTGTACAACCTAACTCACTGG 2167
QY 161 CAAGAACACAGTGTGGGCTTTCGACCACATAGAACAACTTTTCAATTGACAGTTGC 220
Db 2168 CGAGNATACAGCTGGGACCTTCAGCCACTACAAAGAAATTTTAAATTGACAGTTGC 2227
QY 221 AGAATTGTGAGTGTTTTACATTGATCTTTGTAAATGACGTTAGCAGTATGTTTTCGA 280
Db 2228 AGAATTGTGAGTGTTTTACATTGATCTTTGTCTAATGCAATAGCAGTATGTTTTCGA 2287
QY 281 TGTATGACTTAATAATCCTTGAATCAT 308
Db 2288 TGTATGACTTAATAATCCTTGAATCAT 2315

RESULT 13
AX163754
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE
Gene expression profiles in liver cancer
JOURNAL
Patent: WO 0229103-A 1632 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
1..1232
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. D25274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN

Query Match 53.9%; Score 187.4; DB 6; Length 1232;
Best Local Similarity 80.9%; Pred. No. 2.2e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
QY 4 AACATCACTGCTTCCAGAGTACCAACACTGTCATGATTGATGCGCCCTCTAGA 63
Db 905 AACATCACTGCTTCCAGATTACCGACACTGTCATTGACCAATACTGA-CCTCTCTTA 963
QY 64 CCTCACCACGGGACACATGCTTCGGTA-----CCTTTGGGT 102
Db 964 CCTCGCCACGGGACACACGGCTCTGTAGTCGCTTTCCTATTGATGTTCTTTGGGT 1023
QY 103 CTGTGAGGTTCTGTCAA--GCGTAGTGTCTAACGCCGTTCTGTACAACCTAACTCACTGG 160
Db 1024 CTGTGAGGTTCTGTAACTGTCTAGTGTGACGATGTTCTGTACAACCTAACTCACTGG 1083
QY 161 CAAGAACACAGTGTGGGCTTTCGACCACACTAGAACAACTTTTCAATTGACAGTTGC 220
Db 1084 CGAGNATACAGCTGGGACCTTCAGCCACTACAAAGAAATTTTAAATTGACAGTTGC 1143
QY 221 AGAATTGTGAGTGTTTTACATTGATCTTTGTCTAATGCAATGACGATGTTTTCGA 280
Db 1144 AGAATTGTGAGTGTTTTACATTGATCTTTGTCTAATGCAATGACGATGTTTTCGA 1203
QY 281 TGTATGACTTAATAATCCTTGAAT 305
Db 1204 TGTATGACTTAATAATCCTTGGAA 1228

RESULT 15
HUMPO2ST9
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Ikeda, J., Hadano, S., Nagayama, T., Tomiyasu, H., Wakasa, K. and
TITLE
Isolation and characterization of 21 novel expressed DNA sequences
from the distal region of human chromosome 4p

```

```

Db 300 ATCCTGTAATCATACGACTGGTAATACTAGTGTCTTTTGAGACTGTGATGAACA 352

RESULT 14
AX408985
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE
Gene expression profiles in liver cancer
JOURNAL
Patent: WO 0229103-A 1632 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
1..1232
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. D25274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN

Query Match 53.9%; Score 187.4; DB 6; Length 1232;
Best Local Similarity 80.9%; Pred. No. 2.2e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
QY 4 AACATCACTGCTTCCAGAGTACCAACACTGTCATGATTGATGCGCCCTCTAGA 63
Db 905 AACATCACTGCTTCCAGATTACCGACACTGTCATTGACCAATACTGA-CCTCTCTTA 963
QY 64 CCTCACCACGGGACACATGCTTCGGTA-----CCTTTGGGT 102
Db 964 CCTCGCCACGGGACACACGGCTCTGTAGTCGCTTTCCTATTGATGTTCTTTGGGT 1023
QY 103 CTGTGAGGTTCTGTCAA--GCGTAGTGTCTAACGCCGTTCTGTACAACCTAACTCACTGG 160
Db 1024 CTGTGAGGTTCTGTAACTGTCTAGTGTGACGATGTTCTGTACAACCTAACTCACTGG 1083
QY 161 CAAGAACACAGTGTGGGCTTTCGACCACACTAGAACAACTTTTCAATTGACAGTTGC 220
Db 1084 CGAGNATACAGCTGGGACCTTCAGCCACTACAAAGAAATTTTAAATTGACAGTTGC 1143
QY 221 AGAATTGTGAGTGTTTTACATTGATCTTTGTCTAATGCAATGACGATGTTTTCGA 280
Db 1144 AGAATTGTGAGTGTTTTACATTGATCTTTGTCTAATGCAATGACGATGTTTTCGA 1203
QY 281 TGTATGACTTAATAATCCTTGAAT 305
Db 1204 TGTATGACTTAATAATCCTTGGAA 1228

RESULT 15
HUMPO2ST9
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Ikeda, J., Hadano, S., Nagayama, T., Tomiyasu, H., Wakasa, K. and
TITLE
Isolation and characterization of 21 novel expressed DNA sequences
from the distal region of human chromosome 4p

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:50:08 ; Search time 26.5359 Seconds
(without alignments)
5788.425 Million cell updates/sec

Title: US-09-717-321A-17
Perfect score: 348
Sequence: 1 tgaacatcactgtctgcc.....ctttggaacttgaaaaaaa 348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	52.4	15.1	605	3	US-09-385-982-483
C 2	43	12.4	38844	4	US-09-734-675-3
C 3	40.8	11.7	5852	1	US-07-867-106-2
C 4	39.2	11.3	859	1	US-08-345-756-8
C 5	39.2	11.3	859	1	US-08-625-198-8
C 6	39.2	11.3	2039	1	US-08-345-756-5
C 7	39.2	11.3	2039	1	US-08-625-198-5
C 8	38.8	11.1	760	4	US-09-205-258-232
C 9	38.8	11.1	2218	4	US-09-205-258-103
C 10	36.8	10.6	1217	3	US-09-277-716-17
C 11	36.8	10.6	1217	4	US-09-609-161B-17
C 12	36.4	10.5	1454	4	US-08-745-995A-31
C 13	36.4	10.5	1454	4	US-08-745-995A-33
C 14	36.4	10.5	1454	4	US-09-005-352-31
C 15	36.4	10.5	1454	4	US-09-005-352-33
C 16	36.4	10.5	1584	4	US-08-745-995A-1
C 17	36.4	10.5	1584	4	US-08-745-995A-3
C 18	36.4	10.5	1584	4	US-09-005-352-1
C 19	36.4	10.5	1584	4	US-09-005-352-3
C 20	36.4	10.5	1838	4	US-08-745-995A-13
C 21	36.4	10.5	1838	4	US-08-745-995A-15
C 22	36.4	10.5	1838	4	US-09-005-352-13
C 23	36.4	10.5	1838	4	US-09-005-352-15
C 24	36	10.3	1882	3	US-09-370-253-1
C 25	36	10.3	3736	2	US-08-480-473B-1
C 26	36	10.3	3736	3	US-08-915-213-1
C 27	36	10.3	3736	3	US-09-148-547-1

```

28      36      10.3      3736      3      US-09-235-217-1      Sequence 1, Appli
29      36      10.3      3736      4      US-09-383-581-1      Sequence 1, Appli
30      36      10.3      3736      5      PCT-US96-10251-1      Sequence 1, Appli
31      35.8      10.3      566      4      US-09-105-542A-1      Sequence 1, Appli
32      35.6      10.2      501      4      US-09-601-198-170      Sequence 170, App
33      35.6      10.2      991      3      US-08-924-747-25      Sequence 25, Appli
34      35.6      10.2      991      3      US-09-247-373B-25      Sequence 25, Appli
35      35.6      10.2      991      3      US-09-296-715-25      Sequence 19, Appli
36      35.6      10.2      9829      4      US-09-322-478-19      Sequence 19, Appli
37      35.6      10.2      12286      4      US-09-322-478-17      Sequence 17, Appli
38      35.4      10.2      1571      5      PCT-US94-12912-2      Sequence 2, Appli
39      35.2      10.1      615      4      US-09-105-542A-2      Sequence 2, Appli
40      35      10.1      580073      4      US-08-545-528D-1      Sequence 1, Appli
41      34.8      10.0      13865      3      US-09-009-217-11      Sequence 11, Appli
42      34.8      10.0      13865      3      US-09-009-656-11      Sequence 11, Appli
43      34.6      9.9      664      4      US-09-904-615-66      Sequence 66, Appli
44      34.6      9.9      1596      3      US-09-146-950-17      Sequence 17, Appli
45      34.6      9.9      1929      3      US-09-146-950-1      Sequence 1, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-385-982-483/c
; Sequence 483, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: 11
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-483

```

```

Query Match      15.1%; Score 52.4; DB 3; Length 605;
Best Local Similarity 81.1%; Pred. No. 3.7e-05;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

```

```

QY      4      AACATCACTGCTTGGCAGAGTACCAACTGCTGATTGATGCGGCCCTCTAGA 63
Db      162      AACATCACTGCTTGGCAGATTACCCACACTGTCACCTGACCAATACTGA-CCCTCTTTA 104
QY      64      CCTCACCCGCGGACACATGCTTCCGGTA 93
Db      103      CCTCGCCCGCGGACACACGCTCTCTGTA 74

```

```

RESULT 2
US-09-734-675-3
; Sequence 3, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND

```

RESULT 3
 US-07-867-106-2/c
 ; Sequence 2, Application US/07867106
 ; Patent No. 5389526
 ; GENERAL INFORMATION:
 ; APPLICANT: Slade, Martin B
 ; APPLICANT: Chang, Andy C M
 ; APPLICANT: Williams, Keith L
 ; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
 ; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
 ; STREET: One Liberty Place 46th Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/867,106
 ; FILING DATE: 19920625
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PJ 7187
 ; APPLICATION NUMBER: PCT/AU90/00530
 ; FILING DATE: 02-NOV-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Feeney, Joanne Longo
 ; REGISTRATION NUMBER: 35,134
 ; REFERENCE/DOCKET NUMBER: RICE-0002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-568-3100
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5852 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)

Query Match 11.3%; Score 39.2; DB 1; Length 859;
Best Local Similarity 54.1%; Pred. NO. 0.14;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0

QY 197 AAACATTTTTCATTCAGAGTGCAGAAATGTGGAGTGTGTTTACATTCGATCTTTTGCTA 256
 |||||
 Db 240 AATTTTGGTTTGAATTTTGTTCGAATCATGTGGGTTCGATTCAGATTCGAGTTCGGGTA 181
 |||||
 QY 257 ATGCAGTTAGCAGTATGTTTTCGATGATGATTAATAATCCTTGAATCATATAAAAAA 316
 |||||
 Db 180 CCATTTTAATATGATGATTTTCTTAAAAAAATTCAAATATATACTTAAATCGTCAAAATCA 121
 |||||
 QY 317 AAAAAAAATGCTTTGGAACTTGAAAA 344
 |||||
 Db 120 AAAAAATAAATAATAAAACATAAAAA 93
 |||||

RESULT 5
 US-08-625-198-8/c
 ; Sequence 8, Application US/08625198
 ; Patent No. 5756324
 ; GENERAL INFORMATION:
 ; APPLICANT: Baszczynski, Chris
 ; APPLICANT: Barbour, Eric
 ; APPLICANT: Hattori, Jiro
 ; APPLICANT: Miki, Brian
 ; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/625,198
 ; FILING DATE: 01-APR-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/345,756
 ; FILING DATE: 22-NOV-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, STEPHEN A.
 ; REGISTRATION NUMBER: 33229/236/PIHI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136

INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 859 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-625-198-8
 Query Match 11.3%; Score 39.2; DB 1; Length 859;
 Best Local Similarity 54.1%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;
 Matches 80; Conservative 0
 QY 197 AAACATTTTTCATTCAGAGTGCAGAAATGTGGAGTGTGTTTACATTCGATCTTTTGCTA 256
 |||||
 Db 240 AATTTTGGTTTGAATTTTGTTCGAATCATGTGGGTTCGATTCAGATTCGAGTTCGGGTA 181
 |||||
 QY 257 ATGCAGTTAGCAGTATGTTTTCGATGATGATTAATAATCCTTGAATCATATAAAAAA 316
 |||||
 Db 180 CCATTTTAATATGATGATTTTCTTAAAAAAATTCAAATATATACTTAAATCGTCAAAATCA 121
 |||||

RESULT 6
 US-08-345-756-5/c
 ; Sequence 5, Application US/08345756
 ; Patent No. 5633438
 ; GENERAL INFORMATION:
 ; APPLICANT: Baszczynski, Chris
 ; APPLICANT: Barbour, Eric
 ; APPLICANT: Hattori, Jiro
 ; APPLICANT: Miki, Brian
 ; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/345,756
 ; FILING DATE: 22-NOV-1994
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, STEPHEN A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 33229/236/PIHI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2039 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-345-756-5

Query Match 11.3%; Score 39.2; DB 1; Length 2039;
 Best Local Similarity 54.1%; Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;
 Matches 80; Conservative 0
 QY 197 AAACATTTTTCATTCAGAGTGCAGAAATGTGGAGTGTGTTTACATTCGATCTTTTGCTA 256
 |||||
 Db 231 AATTTTGGTTTGAATTTTGTTCGAATCATGTGGGTTCGATTCAGATTCGAGTTCGGGTA 172
 |||||
 QY 257 ATGCAGTTAGCAGTATGTTTTCGATGATGATTAATAATCCTTGAATCATATAAAAAA 316
 |||||
 Db 171 CCATTTTAATATGATGATTTTCTTAAAAAAATTCAAATATATACTTAAATCGTCAAAATCA 112
 |||||
 QY 317 AAAAAAAATGCTTTGGAACTTGAAAA 344
 |||||
 Db 111 AAAAAATAAATAATAAAACATAAAAA 84
 |||||

RESULT 7
 US-08-625-198-5/c
 ; Sequence 5, Application US/08625198
 ; Patent No. 5756324
 ; GENERAL INFORMATION:
 ; APPLICANT: Baszczynski, Chris
 ; APPLICANT: Barbour, Eric

	APPLICANT:	Hattori, Jiro	
	APPLICANT:	Miki, Brian	
	TITLE OF INVENTION:	MICROSPORE-SPECIFIC REGULATORY ELEMENT	
	NUMBER OF SEQUENCES:	8	
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE:	Foley & Lardner	
	STREET:	3000 K Street, N.W., Suite 500	
	CITY:	Washington	
	STATE:	D.C.	
	COUNTRY:	USA	
	ZIP:	20007-5109	
	COMPUTER READABLE FORM:		
	MEDIUM TYPE:	Floppy disk	
	COMPUTER:	IBM PC compatible	
	OPERATING SYSTEM:	PC-DOS/MS-DOS	
	SOFTWARE:	PatentIn Release #1.0, Version #1.30	
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER:	US/08/625,198	
	FILING DATE:	01-APR-1996	
	CLASSIFICATION:	435	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 08/345,756	
	FILING DATE:	22-NOV-1994	
	ATTORNEY/AGENT INFORMATION:		
	NAME:	BENT, STEPHEN A.	
	REGISTRATION NUMBER:	29,768	
	REFERENCE/DOCKET NUMBER:	33229/236/PIHI	
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE:	(202)672-5300	
	TELEFAX:	(202)672-5399	
	TELEX:	904136	
	INFORMATION FOR SEQ ID NO:	5:	
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	2039 base pairs	
	TYPE:	nucleic acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	linear	
	MOLECULE TYPE:	DNA (genomic)	
	US-0625-198-5		
	Query Match	11.3%; Score 39.2; DB 1; Length 2039;	
	Best Local Similarity	54.1%; Pred. No. 0.18;	
	Matches	80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;	
QY		197 AAACATTTTCAATGCAGATTCGAGAATTCGGCGATTTTGACATGTACTTTTGCTA 256	
Db		231 RAATTTGGTTGAAATTTTGTCGAATCATTCGGTTCGATTCAGATTCGAGTCGGTA 172	
QY		257 ATGCAGTAGCAGTAGTATGTTTGCATGTATGACTTAATAAATCCCTTGATCATAAAAAAA 316	
Db		171 CCCATTTTAAATATGCTATTTTTCTTTAAAAAAATCAAATATACATACTTAATCGTCAAAATCA 112	
QY		317 AAAAAAAAATGCTTTTGGAACTTGAAAA 344	
Db		111 AAAAATAAAAATAATAATAAACATAAAAA 84	
RESULT 8			
US-09-205-258-232			
/ Sequence 232, Application US/09205258			
/ Patent No. 6525174			
/ GENERAL INFORMATION:			
/ APPLICANT:	Young et al.		
/ TITLE OF INVENTION:	207 Human Secreted Proteins		
/ FILE REFERENCE:	PZ007P1		
/ CURRENT APPLICATION NUMBER:	US/09/205,258		
/ CURRENT FILING DATE:	1998-12-04		
/ EARLIER APPLICATION NUMBER:	PCT/US98/11422		
/ EARLIER FILING DATE:	1998-06-04		
/ EARLIER APPLICATION NUMBER:	60/048,885		
/ EARLIER FILING DATE:	1997-06-06		
/ EARLIER APPLICATION NUMBER:	60/049,375		
/ EARLIER FILING DATE:	1997-06-06		

```

; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 232
; LENGTH: 760
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (741)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-205-258-232

```

	Query Match	11.1%;	Score 38.8;	DB 4;	Length 760;
	Best Local Similarity	55.1%;	Pred. No. 0.17;		
	Matches	76;	Conservative 0;	Mismatches 62;	Indels 0; Gaps 0;
QY	211	TGACAGTGTGCAGAAATCTGGAGTCTTTTACATTCATCTTTGGCTAATGCGATGTAGCAGT	270		
Db	602	TGAAGCTATGTACTAATCTTAGTTTGTAAATGTCTTTTGATACCATCATCTTCGTTT	661		
QY	271	ATCTTTTGGCATGTATGACTTAAUAAATCCCTTGAATCATATAAAAAAATAAAAAAATGCTCT	330		
Db	662	TCCTTTTGTAGGTATAAATAAACAACATGTTGACAATAAAAAAATAAAAAAATAAAAAA	721		
QY	331	TTTGAACCTTGAAAAAAA	348		
Db	722	AAAAAATAAAAAAATAAAAAA	739		

```

RESULT 9
US-09-205-258-103
; Sequence 103, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: PZ007F1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899

```

```

/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,893
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,900
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,901
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,892
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,915
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,019
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,970
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,972
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,916
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,373
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,875
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,374
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,917
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,949
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,974
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,883
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,897
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,898
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,962
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,963
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,877
/ EARLIER FILING DATE: 1997-12-18
/ EARLIER APPLICATION NUMBER: 60/092,921
/ EARLIER FILING DATE: 1998-07-15
/ EARLIER APPLICATION NUMBER: 60/094,657
/ EARLIER FILING DATE: 1998-07-30
/ NUMBER OF SEQ ID NOS: 1227
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 103
/ LENGTH: 2218
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-0205-258-103

```

	Query Match	11.1%	Score 38.8	DB 4	Length 2218	
	Best Local Similarity	55.1%	Pred. No. 0.24			
	Matches	76	Conservative	0	Mismatches	62
					Indels	0
					Gaps	0
Qy	211	TGACAGTTGCAGAAATTGTGGAGTGTGTTTTTACATTGATCTTTTTCGTAAATGCAGTTAGCAGT	270			
Db	2061	TGAAGAGTATGTTACTATCTTCTTAGTTTGTAAATGTGCTTTTGATACCATCATCTTGTGTTT	2120			
Qy	271	AGTTTTTGGCATGTATGACTTAATAAATCCCTTGAATCATATAAAAAAATAAAAAAATGCT	330			
Db	2121	TCTTTTTGTAGGTATAAATAAAAAACACTGTGACATAAAAAAATAAAAAAATAAAAAA	2180			
Qy	331	TTGGAACTTGAAAAAAA	348			

```
Db      2181 AAAAAAAAAAAAAAAAAA 2198

RESULT 10
US-09-277-716-17
; Sequence 17, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1217
; TYPE: DNA
; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (31)..(963)
; OTHER INFORMATION: Renilla mulleri luciferase
;
US-09-277-716-17
Query Match      10.6%; Score 36.8; DB 3; Length 1217;
Best Local Similarity 63.6%; Pred. No. 0.69;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      238 TTACATGATCTTTGCTAATGACGTAGCAGTATGTTGCATGTATGACTTAATAAT 297
      1126 TTCTCTTAGCTATTATATAATACATCTTATGTAATAAAACCTTGTGTTAATAAT 1185

Db      298 CCTTGATCATAAAAAAAAAAAAAA 325
      1186 TAATGATTCAGAAAAAAAAAAAAA 1213

RESULT 11
US-09-609-161B-17
; Sequence 17, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1217
; TYPE: DNA

; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (31)..(963)
; OTHER INFORMATION: Renilla mulleri luciferase
;
US-09-717-321a-17.rni
Query Match      10.6%; Score 36.8; DB 3; Length 1217;
Best Local Similarity 63.6%; Pred. No. 0.69;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      238 TTACATGATCTTTGCTAATGACGTAGCAGTATGTTGCATGTATGACTTAATAAT 297
      1126 TTCTCTTAGCTATTATATAATACATCTTATGTAATAAAACCTTGTGTTAATAAT 1185

Db      298 CCTTGATCATAAAAAAAAAAAAAA 325
      1186 TAATGATTCAGAAAAAAAAAAAAA 1213

RESULT 12
US-08-745-995A-31
; Sequence 31, Application US/08745995A
; Patent No. 6372887
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary M.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: No. 6372887el Serine Protease Inhibitor
; TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
; TITLE OF INVENTION: and Uses Thereof
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745,995A
; FILING DATE: 07-NOV-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1210
;
US-08-745-995A-31
Query Match      10.5%; Score 36.4; DB 4; Length 1454;
Best Local Similarity 58.2%; Pred. No. 0.93;
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY      216 GTTCAGCAATGTCAGTGTGTTTACATTCATCTTTGCTAATGACGTATGACGACTATGTT 275
      1345 GTAGACGAAATGTTTTGTTTGTAGTTTCATCTTTTATGAATGTAATACCTATAAT 1404

Db
```

QY 276 TTGCATGTATGACTTAATAAATCCTTGTAATCATATAAAAAAAAAAAAAA 325
 Db 1405 GTTGAGTTTATGTTAATAAATGTTAAATGTGAAAAAATAAAAAAAAA 1454

RESULT 13

US-08-745-995A-33/c
 ; Sequence 33, Application US/08745995A
 ; Patent No. 6372887

GENERAL INFORMATION:

APPLICANT: Silver, Gary M.
 APPLICANT: Wisniewski, Nancy
 TITLE OF INVENTION: No. 6372887el Serine Protease Inhibitor
 TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
 TITLE OF INVENTION: and Uses Thereof
 NUMBER OF SEQUENCES: 41
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Heska Corporation
 STREET: 1825 Sharp Point Drive
 CITY: Fort Collins
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: WordPerfect for Windows, Version 7.0
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/745,995A
 FILING DATE: 07-NOV-1996
 CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington
 REGISTRATION NUMBER: 37,459
 REFERENCE/DOCKET NUMBER: FC-2
 TELEPHONE: 970/493-7272
 TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1454 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA

US-08-745-995A-33

Query Match 10.5%; Score 36.4; DB 4; Length 1454;
 Best Local Similarity 58.3%; Pred. No. 0.93;
 Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 216 GTTGACGAATGTGGAGTGTCTTTTACATGATCTTTTGCTTAATGCAGTTAGCAGTATGTT 275
 Db 110 GTAGACGAAATGTTTGTGTTTGTAGTTTTCACCTTTTATGAATGTAATCACCCTATATAAT 51

QY 276 TTGCATGTATGACTTAATAAATCCTTGTAATCATATAAAAAAAAAAAAAA 325
 Db 50 GTTGAGTTTATGTTAATAAATGTTAAATGTGAAAAAATAAAAAAAAA 1

RESULT 14

US-09-005-352-31
 ; Sequence 31, Application US/09005352
 ; Patent No. 6479253

GENERAL INFORMATION:

APPLICANT: Silver, Gary M.
 APPLICANT: Wisniewski, Nancy
 TITLE OF INVENTION: No. 6479253el Serine Protease Inhibitor
 TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
 TITLE OF INVENTION: and Uses Thereof
 NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Heska Corporation
 STREET: 1825 Sharp Point Drive
 CITY: Fort Collins
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80525
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: WordPerfect for Windows, Version 7.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/005,352
 FILING DATE: 07-NOV-1996
 CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
 NAME: Verser, Carol Talkington
 REGISTRATION NUMBER: 37,459
 REFERENCE/DOCKET NUMBER: FC-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 970/493-7272
 TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1454 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 20..1210
 US-09-005-352-31

Query Match 10.5%; Score 36.4; DB 4; Length 1454;
 Best Local Similarity 58.2%; Pred. No. 0.93;
 Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 216 GTTGACGAATGTGGAGTGTCTTTTACATGATCTTTTGCTTAATGCAGTTAGCAGTATGTT 275
 Db 1345 GTAGACGAAATGTTTGTGTTTGTAGTTTTCACCTTTTATGAATGTAATCACCCTATAAT 1404

QY 276 TTGCATGTATGACTTAATAAATCCTTGTAATCATATAAAAAAAAAAAAAA 325
 Db 1405 GTTGAGTTTATGTTAATAAATGTTAAATGTGAAAAAATAAAAAAAAA 1454

RESULT 15

US-09-005-352-33/c
 ; Sequence 33, Application US/09005352
 ; Patent No. 6479253

GENERAL INFORMATION:

APPLICANT: Silver, Gary M.
 APPLICANT: Wisniewski, Nancy
 TITLE OF INVENTION: No. 6479253el Serine Protease Inhibitor
 TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
 TITLE OF INVENTION: and Uses Thereof
 NUMBER OF SEQUENCES: 41
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Heska Corporation
 STREET: 1825 Sharp Point Drive
 CITY: Fort Collins
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: WordPerfect for Windows, Version 7.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/005,352

Mon Nov 24 15:20:12 2003

```

;
; FILING DATE: 07-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: PC-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-09-005-352-33

Query Match      10.5%; Score 36.4; DB 4; Length 1454;
Best Local Similarity 58.2%; Pred. No. 0.93; 46; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY      216 GTTCAGAAATGCGAGTGTGTTTACATTGATCTTTTGCTAATGCAGTTAGCAGTATGTT 275
Db      110 GTAGACGAAAAGTTTGTGTTTGTAGTTTTCACCTTTTATGAATGTAATCACCTATATAAT 51

QY      276 TTGCATGTATGACTTAATAAATCCTTGAAATCAATAAAAAAAAAAAAAAAAAA 325
Db      50 GTTGTAGTTTATGTAATAAATAATGTTAAATGTGAAAAAAAAAAAAAAAAAAAA 1

Search completed: November 23, 2003, 15:53:46
Job time : 28.5359 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 14:02:08 : Search time 123.329 Seconds
(without alignments)
9221.962 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348

Sequence: 1 tgaacatcactgtctgcc.....ctttggaacttgaaaaaaa 348

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	211.6	60.8	2941	14	US-10-198-846-9874
2	209.4	60.2	2051	9	US-09-925-302-255
3	192.4	55.3	1644	14	US-10-153-668-48
4	192.4	55.3	1646	14	US-10-153-668-46
5	187.4	53.9	1232	10	US-09-880-107-1632
6	163.6	47.0	353	10	US-09-960-352-11432
7	116.4	33.4	3740	11	US-09-764-891-9984
8	95.4	27.4	201	10	US-09-920-352-4357
9	52.4	15.1	162	10	US-09-920-300A-1549
10	52.4	15.1	162	12	US-10-099-926-1549
11	52.4	15.1	162	13	US-10-033-528-1549
12	52.4	15.1	477	14	US-10-066-543-3022
13	52.4	15.1	549	14	US-10-066-543-2878
14	52.4	15.1	567	14	US-10-066-543-2364
15	52.4	15.1	594	14	US-10-066-543-2864
16	52.4	15.1	605	11	US-09-871-161-483

17	51.8	14.9	2313	14	US-10-198-846-12919	Sequence 12919, A
18	50.8	14.6	602	10	US-09-878-178-1654	Sequence 1654, A
19	50.8	14.6	602	13	US-10-046-935-1654	Sequence 1654, A
20	50.8	14.6	602	14	US-10-146-502-1654	Sequence 1654, A
21	45	12.9	249	12	US-09-814-353-1384	Sequence 1384, A
22	45	12.9	249	12	US-09-814-353-7747	Sequence 7747, A
23	43.2	12.4	424	10	US-09-814-353-14132	Sequence 14132, A
24	43.2	12.4	424	10	US-09-814-353-14132	Sequence 14132, A
25	43	12.4	38844	13	US-10-060-333-3	Sequence 1218, A
26	42.2	12.1	640681	10	US-09-790-988-1	Sequence 3, Appli
27	42	12.1	15261	11	US-09-764-891-10186	Sequence 1, Appli
28	41.8	12.0	9817	12	US-10-311-455-1341	Sequence 10186, A
29	41.4	11.9	6464	12	US-10-311-455-487	Sequence 1341, A
30	40.6	11.7	336	12	US-09-814-353-4742	Sequence 487, App
31	40.6	11.7	336	12	US-09-814-353-11039	Sequence 4742, App
32	40.4	11.6	469	11	US-09-918-995-13017	Sequence 11039, A
33	39.8	11.4	3673778	12	US-10-312-841-1	Sequence 13017, A
34	39.4	11.3	6059	12	US-10-311-455-786	Sequence 1, Appli
35	39.2	11.3	277	10	US-09-960-352-12673	Sequence 786, App
36	39.2	11.3	375	10	US-09-960-352-15014	Sequence 12673, A
37	39.2	11.3	6621	12	US-10-311-455-886	Sequence 15014, A
38	39	11.2	501	12	US-10-027-632-184148	Sequence 886, App
39	39	11.2	501	13	US-10-027-632-184148	Sequence 184148, A
40	39	11.2	1460	14	US-10-198-846-9893	Sequence 184148, A
41	38.8	11.1	760	12	US-09-933-767-103	Sequence 9893, App
42	38.8	11.1	760	14	US-10-023-282-232	Sequence 103, App
43	38.8	11.1	2218	12	US-09-933-767-232	Sequence 232, App
44	38.8	11.1	2218	14	US-10-023-282-103	Sequence 103, App
45	38.8	11.1	2261	11	US-09-764-891-10185	Sequence 10185, A

ALIGNMENTS

RESULT 1

US-10-198-846-9874
; Sequence 9874, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9874
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-9874

Query Match 60.8%; Score 211.6; DB 14; Length 2941;
Best Local Similarity 80.2%; Pred. No. 1.6e-45;
Matches 295; Conservative 0; Mismatches 49; Indels 24; Gaps 3;

QY	4	AACATCAGTCTTGGCCAGGTACCAACACTGTCATGTCATTCGCGTA	-----CCTTTGGGT 102
Db	2106	AAATCAGTCTTGGCCAGGTACCAACACTGTCATGTCATTCGCGTA	-----CCTTTGGGT 102
QY	64	CTCACCCACGGGACACATGTTCCGTA	-----CCTTTGGGT 102
Db	2165	CCTCGCCACGGGACACATGTTCCGTA	-----CCTTTGGGT 102
QY	103	CTGTAGGTTCTGTCAA--GCCTAGTGTACGGCGTCTGTACAACCTACTACTG 160	-----CCTTTGGGT 102

Db	2225	CTGTGAGGTTCTGTAAACTGTGTAGTGTGACGATGTTCTGTACAACTTAACTCACTGG	2284
Qy	161	CAAGAACACAGTGTGGCCCTTTGCACCACTAGAACAAACTTTTTTCAATTGACAGTTGC	220
Db	2285	CGAAGATACAGCTGGGACCCCTTCAGCCACTACACAGAATTTTTTAAATTGACAGTTGC	2344
Qy	221	AGAATTGGGAGTGTTTTTACATTGATCTTTTTGCTAATGCAGTTAGCAGTATGTTTTGCA	280
Db	2345	AGAATTGTGGAGTCTTTTTTACATTGATCTTTTTGCTAATGCAATTAGCATTTATGTTTTGCA	2404
Qy	281	TGTTATGCTTAATAATAATCCCTTGAATCATAAAAAATAAAAAAATCTCTTGGAACTTG	340
Db	2405	TGTTATGCTTAATAATAATCCCTTGAATCATAAAAAATAAAAAAATAAAAAAATAAAAA	2464
Qy	341	AAAAAAAA 348	
Db	2465	AAAAAAAA 2472	

```

RESULT 2
US-09-925-302-255
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-255

```

Qy 161 CAGAACACAGTGTGGCCCTTCGACCACATAGACAACTTTTCAATTGACAGTTGC 220

Db 1845 CGAATATACAGCGTGGGACCCCTTCAGCCACTACACAGAAATTTTTTAAATTTGACAGTTGC 1904

Qy 221 AGAATTCGGAGTGTTTTTTACATTGATCTTTTCTGAATGCAGTTAGCAGTATCTTTTGCA 280

Db 1905 AGAATTCGGAGTGTTTTTTACATTGATCTTTTCTGAATGCAGTATCTTTTGCA 1964

Qy 281 TGTATGACTTAAATAAATCCTTGAATCATATAAAAAAAAAAAAAAAAAAAAA 325

Db 1965 TGTATGACTTAAATAAATCCTTGAATCATATAAAAAAAAAAAAAAAAAAAAA 2009

RESULT 3

US-10-153-668-48

; Sequence 48, Application US/10153668

; Publication No. US20030092616A1

; GENERAL INFORMATION:

; APPLICANT: HONDA, Goichi

; APPLICANT: MATSUDA, Akio

; APPLICANT: MURAMATSU, Shuji

; APPLICANT: ISHIZAWA, Kenya

; TITLE OF INVENTION: STAT6 Activating Gene

; FILE REFERENCE: 1254-0207P

; CURRENT APPLICATION NUMBER: US/10/153,668

; CURRENT FILING DATE: 2002-05-24

; PRIOR APPLICATION NUMBER: US 60/293,172

; PRIOR FILING DATE: 2001-05-25

; PRIOR APPLICATION NUMBER: US 60/316,031

; PRIOR FILING DATE: 2001-08-31

; PRIOR APPLICATION NUMBER: US 60/328,403

; PRIOR FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: JP 2001-157043

; PRIOR FILING DATE: 2001-05-25

; PRIOR APPLICATION NUMBER: JP 2001-260681

; PRIOR FILING DATE: 2001-08-30

; PRIOR APPLICATION NUMBER: JP 2001-313175

; PRIOR FILING DATE: 2001-10-10

; NUMBER OF SEQ ID NOS: 488

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 48

; LENGTH: 1644

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (109)..(447)

US-10-153-668-48

Query Match	55.3%	Score 192.4	DB 14	Length 1644
Best Local Similarity	81.7%	Mism. No. 1.4e-40		
Matches 268	Conservative	0	Mismatches 36	Indels 24
				Gaps 3
QY	4	RACATCACTGCTTCGCCAGAGTACCAACACTGTGATGATGATGCGCCGCCCTCTAGA	63	
Db	1318	AACTATCATGCTCTGCCAGATTACGACACTGTCACTGACCAATACTGA-CCCTCTTTA	1376	
QY	64	CCTCACCCACGCGGACACATGTTCCGGTA-----CCTTTGGGT	102	
Db	1377	CCTCGCCACGCGGACACACGCCTCTCTAGTCGTTTGGCTATTGATGTTCCTTGGGT	1436	
QY	103	CTGTGAGGTTCTGTCAA--GGCTAGTGTAAACGCGTTCGTGTACAACTAACTCACTGG	160	
Db	1437	CTGTGAGTTCTGTAACTGTGCTAGTCTGCGGATGTTCTGTACAACTTAACTCACTGG	1496	
QY	161	CAAGAACACAGTGTGGGCTTTGACCACTAGAACAACTTTTTCATTTGACAGTTGC	220	
Db	1497	CGAGATAACGCGTGGGACCTTCAGCCACTACAACAAATTTTTTAAATTGACAGTTGC	155	
QY	221	AGAATTGGGAGTGTTTTTACATTGATCTTTGCTTAATGCACTTACAGTATGTTTTGCA	280	
Db	1557	AGAAATTGGAGTGTTTTTTACATTGATCTTTGCTTAATGCACTTACAGTATGTTTTGCA	1611	

QY 281 TGTATGACTTAATAAATCCTTGAATCAT 308
|||||
Db 1617 TGTATGACTTAATAAATCCTTGAATCAT 1644

RESULT 4

US-10-153-668-46
; Sequence 46, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT5 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(273)
US-10-153-668-46

Query Match 55.3%; Score 192.4; DB 14; Length 1646;
Best Local Similarity 81.7%; Pred. No. 1.4e-40;
Matches 268; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

QY 4 AACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCTAGA 63
|||||
Db 1320 AACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCTAGA 1378
|||||
QY 64 CCTCACCCACGGGACACATGCTTCCGGTA-----CCTTTGGGT 102
|||||
Db 1379 CCTCGCCACGGGACACACGCTCTCTGTAGTCGCTTTGGCTATTGATGTTCCTTGGGT 1438
|||||
QY 103 CTGTGAGGTTCTGTCAA--GCGTAGTGCTAAAGCCCTTCTGTACAACTAACTCACTGG 160
|||||
Db 1439 CTGTGAGGTTCTGTAAACTGTCTAGTGTGAGAGTGTCTGTACAACTAACTCACTGG 1498
|||||
QY 161 CAGACACAGTGTGGGCTTTCGACCACTAGAAACAACTTTTCAATGACAGTTGC 220
|||||
Db 1499 CGAATACAGGTGGGACCTTCAGCCACTCAACAGAAATTTTAAATGACAGTTGC 1558
|||||
QY 221 AGAATGTGAGTGTTTACATTTGATGATCTTTGCTAATGAGTACAGTATGTTTGA 280
|||||
Db 1559 AGAATGTGAGTGTTTACATTTGATGATCTTTGCTAATGCAATAGCAATTATGTTTGA 1618
|||||
QY 281 TGTATGACTTAATAAATCCTTGAATCAT 308
|||||
Db 1619 TGTATGACTTAATAAATCCTTGAATCAT 1646
|||||

RESULT 5

US-09-880-107-1632
; Sequence 1632, Application US/09880107

; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1632
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D25274
US-09-880-107-1632

Query Match 53.9%; Score 187.4; DB 10; Length 1232;
Best Local Similarity 80.9%; Pred. No. 2.6e-39;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;

QY 4 AACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCTAGA 63
|||||
Db 905 AACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCTAGA 963
|||||
QY 64 CCTCACCCACGGGACACATGCTTCCGGTA-----CCTTTGGGT 102
|||||
Db 964 CCTCGCCACGGGACACACGCTCTCTGTAGTCGCTTTGGCTATTGATGTTCCTTGGGT 1023
|||||
QY 103 CTGTGAGGTTCTGTCAA--GCGTAGTGCTAAAGCCCTTCTGTACAACTAACTCACTGG 160
|||||
Db 1024 CTGTGAGGTTCTGTAAACTGTCTAGTGTGACGATGTTCTGTACAACTAACTCACTGG 1083
|||||
QY 161 CAAGACACAGTGTGGGCTTTCGACCACTAGAACAACTTTTCAATGACAGTTGC 220
|||||
Db 1084 CGAATACAGGTGGGACCTTCAGCCACTCAACAGAAATTTTAAATGACAGTTGC 1143
|||||
QY 221 AGAATGTGAGTGTTTACATTTGATGATCTTTGCTAATGAGTACAGTATGTTTGA 280
|||||
Db 1144 AGAATGTGAGTGTTTACATTTGATGATCTTTGCTAATGCAATAGCAATTATGTTTGA 1203
|||||
QY 281 TGTATGACTTAATAAATCCTTGAAT 305
|||||
Db 1204 TGTATGACTTAATAAATCCTTGGAA 1228
|||||

RESULT 6

US-09-960-352-11432/c
; Sequence 11432, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11432
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 49-L1B3058-009-Q1-K1-E2

US-09-960-352-11432

	Query Match	47.0%; Score 163.6; DB 10; Length 353;
	Best Local Similarity	90.3%; Pred. No. 2.8e-33;
	Matches 186; Conservative	0; Mismatches 19; Indels 1; Gaps 1;
Qy	120	GCGCTAGTGTAAACGCCGTTCTGTACAAACCTTAACACTCTGGCAAGAACAAGTGTTGGGC 179
Db	246	GTGCTGTGCTAACGAAGTTCGTACAA-CITACCACCTGSCAAGATAACGAGCTTGGAC 188
Qy	180	CTTTGCAACCATAGAAACAACTTTTTTCAATTGACAGTTGCAGAAATTTGTGGAGTGTTTT 239
Db	187	CTTTTCAAACCATAGAAACAAAATTTTTTAAATTTGACAGTTGCAGAAATTTGTGGAGTGTTTT 128
Qy	240	ACATTGATCTTTTGCCTAATGCAGTTAGCAGTAGTCTTTTGCATGATGATGACTTAATAAATCC 299
Db	127	ACATTGATCTTTTGCCTACGCCATTAGCAGTAGTCTTTTGCATGATGATGACTTAATAAACCC 68
Qy	300	TTGAATCATAAAAAIAAAAAAAAAAAAA 325
Db	67	TTGAATCATAAAAAIAAAAAAAAAAAAA 42

RESULT 7

```

US-09-764-891-9984
; Sequence 9984, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9984
; LENGTH: 3740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9984

```

RESULT 8

```

US-09-960-352-4357/c
; Sequence 4357, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24

```

; NUMBER OF SEQ ID NOS: 151112

```
; SEQ ID NO 4357
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 19-LIB3058-009-QI-K1-E3
US-09-960-352-4357

Query Match          27.4%; Score 95.4; DB 10; Length 201;
Best Local Similarity 76.5%; Pred. No. 2.1e-15;
Matches 117; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY      174   TTGGGCGCTTTCGCACCACATAGAACAAACTTTTTTTCAATTGCAGGTGCAGAATTGTGGAGT 233
        |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db       193   TTCGAACCTTTTCAAATAATTAGAACAATAATTTTTTAATTTGACAGTTGCAGAATTTTGAGT 134

QY      234   GTTTTTACATGTGATCTTTTCCTAATGCAGTTAGCAGTATCTTTTGCATGTATGACTTTAAT 293
        |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db       133   GAATTTAGATTGATTTTTCATACGTGATAAAAGTATCTTTGAAGGTATCCCTTAAT 74

QY      294   AAATCCCTTGATCATAAAAAANAANAAAAAAAAAT 326
        |||    |||    |||    |||    |||    |||    |||    |||    |||    |||

Gb       73   AATTCTTTGCCCTCAGTATAAGATAAAAT 41
```

RESULT 9

```

US-09-920-300A-1549
; Sequence 1549, Application US/09920300A
; Patent No. US2002036728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: COMPOSITIONS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1549
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-920-300A-1549

```

RESULT 10

```

RESUBM 10
US-10-099-926-1549
; Sequence 1549, Application US/10099926
; Publication No. US2003016064A1
;
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF COLON CANCER
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C2

```

```

; CURRENT APPLICATION NUMBER: US/10/099,926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1549
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-099-926-1549

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 12; Length 162;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGCTTCCAGAGTACCAACACTGTCATGATGATGCGGCCCTCTAGA 63
Db 39 AACATCACTGCTTCCAGATTACCGACACTGTCATGACCAATACTGA-CCCTCTTTA 97

QY 64 CCTCACCACGGGACACATGCTTCCGTA 93
Db 98 CCTCGCCACGGGACACACGGCTCTCTGA 127

RESULT 11
US-10-033-528-1549
; Sequence 1549, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033.528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1549
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1549

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 13; Length 162;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGCTTCCAGAGTACCAACACTGTCATGATGATGCGGCCCTCTAGA 63
Db 39 AACATCACTGCTTCCAGATTACCGACACTGTCATGACCAATACTGA-CCCTCTTTA 97

QY 64 CCTCACCACGGGACACATGCTTCCGTA 93
Db 98 CCTCGCCACGGGACACACGGCTCTCTGA 127

RESULT 12
US-10-066-543-3022/c
; Sequence 3022, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita

```

```

; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-3022

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 14; Length 477;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGCTTCCAGAGTACCAACACTGTCATGATGATGCGGCCCTCTAGA 63
Db 162 AACATCACTGCTTCCAGATTACCGACACTGTCATGACCAATACTGA-CCCTCTTTA 104

QY 64 CCTCACCACGGGACACATGCTTCCGTA 93
Db 103 CCTCGCCACGGGACACACGGCTCTCTGA 74

RESULT 13
US-10-066-543-2878/c
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 526
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2878

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 14; Length 549;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGCTTCCAGAGTACCAACACTGTCATGATGATGCGGCCCTCTAGA 63
Db 162 AACATCACTGCTTCCAGATTACCGACACTGTCATGACCAATACTGA-CCCTCTTTA 104

QY 64 CCTCACCACGGGACACATGCTTCCGTA 93
Db 103 CCTCGCCACGGGACACACGGCTCTCTGA 74

```

RESULT 14

US-10-066-543-2964/c
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yudiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2964

Query Match 15.1%; Score 52.4; DB 14; Length 567;
Best Local Similarity 81.1%; Pred. No. 0.00074;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGTCTTCCAGAGTACCAACACTGTCTATGTGATGCGGCCCTCTAGA 63
DB 162 AACATCACTGTCTTGCAGATTACCGACACTGTGACTTGACCAATACTGA-CCCTCTTTA 104

QY 64 CCTCACCCACGGGACACATGCTTCGGTA 93
DB 103 CCTGCCCCACGGGACACACGCTCTCTGTA 74

RESULT 15

US-10-066-543-2864
; Sequence 2864, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yudiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2864
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-066-543-2864

Query Match 15.1%; Score 52.4; DB 14; Length 594;
Best Local Similarity 81.1%; Pred. No. 0.00076;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGTCTTCCAGAGTACCAACACTGTCTATGTGATGCGGCCCTCTAGA 63
DB 436 AACATCACTGTCTTCCAGATTACCGACACTGTGACTTGACCAATACTGA-CCCTCTTTA 494

QY 64 CCTCACCCACGGGACACATGCTTCGGTA 93
DB 495 CCTGCCCCACGGGACACACGCTCTCTGTA 524

Search completed: November 23, 2003, 19:05:55
Job time : 127.329 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:18:57 ; Search time 460.915 Seconds
(without alignments)
10917.162 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caattgaaaaagtgtgttc.....ctcacagacccaaaggtacc 123

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	123	100.0	123	6	AX163737	AX163737 Sequence
2	123	100.0	348	6	AX163753	AX163753 Sequence
3	123	100.0	1266	6	AX163738	AX163738 Sequence
4	122	99.2	217700	2	AC106124	AC106124 Rattus no
5	107.6	87.5	2281	10	BC051053	BC051053 Mus muscu
6	107.6	87.5	2319	10	BC003828	BC003828 Mus muscu
7	107.6	87.5	269081	2	AC068493	AC068493 Mus muscu
8	106	86.2	192498	2	AC105979	AC105979 Mus muscu
9	106	86.2	230015	2	AC132602	AC132602 Mus muscu
10	100.2	81.5	238121	2	AC133022	AC133022 Rattus no
11	100.2	81.5	239768	2	AC112582	AC112582 Rattus no
12	100.2	81.5	245468	2	AC130746	AC130746 Rattus no
13	87.2	70.9	240973	2	AC123247	AC123247 Rattus no
14	87.2	70.9	241048	2	AC129824	AC129824 Rattus no
15	79.6	64.7	352	6	AX163754	AX163754 Sequence
16	79.6	64.7	455	11	G28995	G28995 human STS S
17	79.6	64.7	1232	6	AX408985	AX408985 Sequence
18	79.6	64.7	1232	9	HUMPO28T9	D25274 Homo sapien
19	79.6	64.7	2302	9	BC050687	BC050687 Homo sapi
20	79.6	64.7	2315	9	AK054993	AK054993 Homo sapi
21	79.6	64.7	28567	9	HSAL32695	HSAL32695 Homo sapi
22	79.6	64.7	212827	9	AC009412	AC009412 Homo sapi
23	77.6	63.1	262	11	G31709	G31709 SWS2233 Er
24	74.8	60.8	5544	9	AF542527	AF542527 Homo sapi
25	74.8	60.8	87616	2	AC139405	AC139405 Homo sapi
26	74.8	60.8	137625	9	AC104663	AC104663 Homo sapi
27	50	40.7	170839	2	AC133467	AC133467 Mus muscu
28	48	39.0	101584	9	CNS01D85	AL121655 BAC seque
29	48	39.0	155943	9	AC012364	AC012364 Homo sapi
30	38.2	31.1	110816	9	AC002404	AC002404 Human Chr
31	33.2	27.0	1485	5	AF029250	AF029250 Danio rer
32	33.2	27.0	1647	5	BC042319	BC042319 Danio rer
33	33.2	27.0	142171	2	EX248228	EX248228 Danio rer
34	33.2	27.0	142171	2	EX248228	EX248228 Danio rer
35	33.2	27.0	223350	2	EX465853	EX465853 Danio rer
36	33.2	27.0	223350	2	EX465853	EX465853 Danio rer
37	33	26.8	165551	2	AC106841	AC106841 Mus muscu
38	32.4	26.3	152747	9	AC092576	AC092576 Homo sapi
39	32.4	26.3	237985	2	AC125754	AC125754 Rattus no
40	32.4	26.3	247478	2	AC097964	AC097964 Rattus no
41	31.4	25.5	605	6	AX341407	AX341407 Sequence
42	31.2	25.4	255150	2	AC103167	AC103167 Rattus no
43	31	25.2	174082	2	AC116798	AC116798 Mus muscu
44	31	25.2	224777	2	AC138109	AC138109 Mus muscu
45	30.8	25.0	156879	10	AL626786	AL626786 Mouse DNA

ALIGNMENTS

RESULT 1
AX163737
LOCUS AX163737
DEFINITION Sequence 1 from Patent WO0138579.
ACCESSION AX163737
VERSION AX163737.1 GI:14544843
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1
REFERENCE
AUTHORS Gould-Rothberg,B.E., Dipippo,V.A., Ramsah,T.M. and Gerwein,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential

Pred. No. is the number of results predicted by chance to have a

RESULT 3
AX163738

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Huiy,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,D., Kowar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,M., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakoelamleh,O., Okwuonu,G., Olarpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfrankoch,C., Plopper,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villalana,D., Waidron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczkyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhauser,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G., and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 217700)
Worley,K.C.

Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217700)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:25094619.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHUL
Center clone name: CH230-119E11
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185725 bases at least Q40
Consensus quality: 190140 bases at least Q30
Consensus quality: 192842 bases at least Q20
Estimated insert size: 194758; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 109797: contig of 109797 bp in length
* 109798 109997: gap of unknown length
* 109998 166934: contig of 57037 bp in length
* 166935 167034: gap of unknown length
* 167035 176384: contig of 9350 bp in length
* 176385 176484: gap of unknown length
* 176485 206327: contig of 29843 bp in length
* 206328 206427: gap of unknown length
* 206428 207461: contig of 1034 bp in length
* 207462 207561: gap of unknown length
* 207562 208607: contig of 1046 bp in length
* 208608 208707: gap of unknown length
* 208708 209908: contig of 1201 bp in length
* 209909 210008: gap of unknown length
* 210009 211683: contig of 1575 bp in length
* 211684 213649: contig of unknown length
* 213650 213749: gap of unknown length
* 213750 215115: contig of 1366 bp in length
* 215116 215216: gap of unknown length
* 215216 217700: contig of 2485 bp in length.

FEATURES

source

1. 217700
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-119E11"

BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others
ORIGIN

Query Match

Best Local Similarity 99.2%; Score 122; DB 2; Length 217700;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CAATTGAAAGTGTCTTCTAGTGTGCGAAGGCCCAACACTGTGTCTTCCAGTGAG	60
Db	64562	CAATTGAAAGTGTCTTCTAGTGTGCGAAGGCCCAACACTGTGTCTTCCAGTGAG	64503
QY	61	TTAGTGTGTACAGACGGCGTTAGCAGCTTGACAGACCTCAGACACCAAGGT	120
Db	64502	TTAGTGTGTACAGACGGCGTTAGCAGCTTGACAGACCTCAGACACCAAGGT	64443
QY	121	AC 122	
Db	64442	AC 64441	

RESULT 5

BC051053/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

BC051053
Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA linear ROD 14-APR-2003
Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.

BC051053.1 GI:29835221

MGC.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2281)

Strausberg,R.

Direct Submission

Submitted (11-APR-2003) National Institutes of Health, Mammalian

REMARK	Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	BC003828/c	BC003828	2319 bp	mRNA	linear	ROD 16-APR-2003
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabbs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nigr.nih.gov Akhter, N., Ayle, K., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, O.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.	BC003828.1	GI:13277917	Mus musculus (house mouse)	Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Plate: 108 Row: f Column: 9 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction, Similarity but not identity to protein.	ACCESSION	BC003828				
	Location/Qualifiers	VERSION	BC003828				
FEATURES	1..2281	KEYWORDS	MGC.				
source	/organism="Mus musculus"	SOURCE	Mus musculus				
	/mol_type="mRNA"	ORGANISM					
	/strain="CZECH II"	REFERENCE					
	/db_xref="taxon:10090"	AUTHORS					
	/clone="MGC:58966 IMAGE:5038182"	JOURNAL					
	/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	MEDLINE	22388257				
	/clone_lib="NCI CGAP_Lu29"	PUBMED	12477932				
	/lab_host="DH10B"	REFERENCE	2 (bases 1 to 2319)				
	/note="Vector: pCMV-SPORT6"	TITLE	Direct Submission				
	197..775	JOURNAL	Strausberg, R.				
	/product="Unknown (protein for MGC:58966)"		Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
CDS	/codon_start=1	REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
	/protein_id="AAH51053.1"	COMMENT	Contact: MGC help desk				
	/db_xref="GI:29835222"		Email: cgabbs-remail.nih.gov				
	/translation="MQAIVGVGAGVGTCTLLISVYTNAPFGEVLPVFDNYSANV MVDKPVNLGLWDAGQEDYDLRLPLSPQTDVFLICFSLVSPASPNVRAKWPYEV RHCPTPIILVGLKDLRDDKDTIEKLKKLTPITYPQGLAWAKEIGAVKYLECSAL TORGLKTVFDEARAVLCPPPVKKRKCKLL"		Tissue Procurement: Gilbert Smith, Ph.D.				
BASE COUNT	568 a 569 c 504 g 640 t		cDNA Library Preparation: Life Technologies, Inc.				
ORIGIN			cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
Query Match	87.5%; Score 107.6; DB 10; Length 2281;		DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;				
Best Local Similarity	92.6%; Pred. No. 1.1e-26;		Web site: http://www.nisc.nih.gov/				
Matches 113; Conservative	0; Mismatches		Contact: nisc_mgc@nigr.nih.gov				
	0; Gaps		0; Indels				
QY	1 CAATTGAAAGAGTTGTTCTAGTGGTCGAAGGCCCAACACTGCTGTTCTTGCCAGTGAG 60						
Db	2170 CAATTGAAAGAGTTGTTCTAGTGGTTGAAGGCCCAACACTGCATCTTGCCAGTGAG 2111						
QY	61 TTAGTTGTACAGAACGCGGTTAGCACTAGCGTTGACAGAACTCAGACACCCAAAGGT 120						
Db	2110 TTAGTTGTACAGACACGCGTTAGCACTAGCACTAGCACTCAGAACCTCAGACACCCCAAGGA 2051						
QY	121 AC 122						
Db	2050 AC 2049						
FEATURES	1..2319	source					
	/organism="Mus musculus"						
	/mol_type="mRNA"						
	/strain="FVB/N"						
	/db_xref="taxon:10090"						
	/clone="MGC:6235 IMAGE:3593957"						
	/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."						
	/clone_lib="NCI CGAP_Maml"						
	/lab_host="DH10B"						
	/note="Vector: pCMV-SPORT6"						
	1..2319						
	/gene="Racl"						
	/note="synonym: D5Erd559e"						
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Plate: 11 Row: e Column: 22. Location/Qualifiers							

```
/db_xref="LocusID:19353"
/db_xref="MGI:97845"
198-776
/codon_start=1
/product="rac1 protein"
/protein_id="AAH03828.1"
/db_xref="GI:13277918"
/db_xref="LocusID:19353"
/translation="MQAICVVGDDGAVGKTCILLISYTNAPGCEYIPTVFDNYSANV
MVGKPNVLGMDTAGEDDRKPLSYPTDPLFCISLVSPASFENVRKAYPEVR
HHCPTDILIVGKLDLRDDEKDTLEKLEKELTPITYPQGLAMAKEIGAVKYLECSAL
TORGLKTVFDEAIRAVLCPFPVKKRKKCLLL"
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN
Query Match 87.5%; Score 107.6; DB 10; Length 2319;
Best Local Similarity 92.6%; Pred. No. 1.1e-26;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 2173 CANTTGAAAAAGTTTCTTCTAGTGGTTGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 2114
QY 61 TTAGTGTGTACGAAAGCGGCTTAGCCTAGCTGCTTGACAGAACTCAGACCCCAAGGT 120
Db 2113 TTAGTGTGTACGAGCAGCGTTAGCCTAGCCTAGCAGTGTGACAGAACTCAGACCCCAAGGA 2054
QY 121 AC 122
Db 2053 AC 2052
RESULT 7
AC068493/c
LOCUS
DEFINITION Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
ACCESSION AC068493
VERSION 1
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 269081)
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 269081)
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
Direct Submission
TITLE Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
JOURNAL Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Brux, NY 10461, USA
COMMENT On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcgg.org/Sequence/mouse.html
Contact: hpcgcmendel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-pp - N/A
**Estimated insert size: 268581 - sum-of-contigs
```

Quality coverage: agarose-pp - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 51730: contig of 51730 bp in length
* 51731: gap of unknown length
* 82423: contig of 30673 bp in length
* 82424: gap of unknown length
* 82444: 118158: contig of 35715 bp in length
* 118159: 118178: gap of unknown length
* 118179: 139441: contig of 21263 bp in length
* 139442: 139461: gap of unknown length
* 139462: 160638: contig of 21177 bp in length
* 160639: gap of unknown length
* 160659: 177611: contig of 16953 bp in length
* 177612: 177631: gap of unknown length
* 177632: 190991: contig of 13360 bp in length
* 190992: 191011: gap of unknown length
* 191012: 200128: contig of 9117 bp in length
* 200129: 200148: gap of unknown length
* 200149: 211882: contig of 11734 bp in length
* 211883: 211902: gap of unknown length
* 211903: 220773: contig of 8871 bp in length
* 220774: 229643: gap of unknown length
* 229644: 229663: gap of unknown length
* 229664: 238339: contig of 8676 bp in length
* 238340: 238359: gap of unknown length
* 238360: 244868: contig of 6489 bp in length
* 244869: 249725: contig of 4857 bp in length
* 249726: 249745: gap of unknown length
* 249746: 252523: contig of 2778 bp in length
* 252524: 252543: gap of unknown length
* 252544: 257520: contig of 4977 bp in length
* 257521: 257540: gap of unknown length
* 257541: 260396: contig of 2856 bp in length
* 260397: 260416: gap of unknown length
* 260417: 262709: contig of 2293 bp in length
* 262710: 262729: gap of unknown length
* 262730: 263934: contig of 1205 bp in length
* 263935: 263954: gap of unknown length
* 263955: 264196: contig of 242 bp in length
* 264197: 264216: gap of unknown length
* 264217: 264629: contig of 413 bp in length
* 264630: 264649: gap of unknown length
* 264650: 265608: contig of 959 bp in length
* 265609: 265628: gap of unknown length
* 265629: 266935: contig of 1307 bp in length
* 266936: 266955: gap of unknown length
* 266956: 268335: contig of 1380 bp in length
* 268336: 268355: gap of unknown length
* 268356: 268549: contig of 194 bp in length
* 268550: 268569: gap of unknown length
* 268570: 269081: contig of 512 bp in length.

FEATURES

Location/Qualifiers
1. 269081
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/clone="RP23-76K1"
/sex="male"
1..51730
/note="assembly_name:Contig138"
51751..82423
misc_feature
misc_feature

```

/misc_feature /note="assembly_name:Contig137" 82444..118158
/misc_feature /note="assembly_name:Contig136" 118179..139441
/misc_feature /note="assembly_name:Contig135" 139462..160638
/misc_feature /note="assembly_name:Contig134" 160859..177611
/misc_feature /note="assembly_name:Contig133" 177632..190991
/misc_feature /note="assembly_name:Contig132" 191012..200128
/misc_feature /note="assembly_name:Contig131" 200149..211862
/misc_feature /note="assembly_name:Contig130" 211903..220773
/misc_feature /note="assembly_name:Contig129" 220794..229643
/misc_feature /note="assembly_name:Contig128" 229664..238339
/misc_feature /note="assembly_name:Contig127" 238360..244848
/misc_feature /note="assembly_name:Contig126" 244869..249725
/misc_feature /note="assembly_name:Contig125" 249746..252523
/misc_feature /note="assembly_name:Contig124 clone end:SP6 vector side:left" 252544..257520
/misc_feature /note="assembly_name:Contig123" 257541..260396
/misc_feature /note="assembly_name:Contig122" 260417..262709
/misc_feature /note="assembly_name:Contig121" 262730..263934
/misc_feature /note="assembly_name:Contig120" 263955..264196
/misc_feature /note="assembly_name:Contig119" 264217..264629
/misc_feature /note="assembly_name:Contig118" 264650..265608
/misc_feature /note="assembly_name:Contig117" 265629..266935
/misc_feature /note="assembly_name:Contig116" 266956..268335
/misc_feature /note="assembly_name:Contig115" 268356..268549
/misc_feature /note="assembly_name:Contig114" 268570..269081
/misc_feature /note="assembly_name:Contig113" 69726 a 66027 c 64672 g 67724 t 932 others
ORIGIN
Query Match 87.5% Score 107.6; DB 2; Length 269081;
Best Local Similarity 92.6%; Pred No. 1.6e-26;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 63360 CAAATTGAAAAAGTTTGTCTAGTGGTTGAAGGCCCAACACTGTGTTCTTGCAGTGAG 63301
QY 61 TTAGGTTGTACAGACGGGTTAGCACTAGCGTTGCAGAACTTCACAGACCCCAAGGT 120
Db 63300 TTAGGTTGTACAGACGGGTTAGCACTAGCGTTGCAGAACTTCACAGACCCCAAGGA 63241
QY 121 AC 122
Db 63240 AC 63239

```

RESULT 8
AC105979

LOCUS
DEFINITION
AC105979
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC105979 192498 bp DNA linear HTG 21-OCT-2002
Mus musculus clone RP24-301E12, WORKING DRAFT SEQUENCE, 9 ordered
pieces.
AC105979
HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192498)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-301E12
Unpublished
2 (bases 1 to 192498)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckghalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,K., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192498)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Bouckghalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,
Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivier,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 21, 2002 this sequence version replaced g1:20149424.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
----- Project name: L20199

```

Center clone name: 301_E_12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190070 bases at least Q40
Consensus quality: 191243 bases at least Q30
Consensus quality: 191514 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 191698; sum-of-contigs
Quality coverage: 11.6 in Q20 bases; agarose-fp
Quality coverage: 9.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1061: contig of 1061 bp in length
* 1062 1161: gap of 100 bp
* 1182 5714: contig of 4553 bp in length
* 5715 5814: gap of 100 bp
* 5815 11422: contig of 5608 bp in length
* 11423 11522: gap of 100 bp
* 11523 17600: contig of 6078 bp in length
* 17601 17700: gap of 100 bp
* 17701 30157: contig of 12457 bp in length
* 30158 30257: gap of 100 bp
* 30258 38787: contig of 8530 bp in length
* 38788 38887: gap of 100 bp
* 38888 99444: contig of 60557 bp in length
* 99445 99544: gap of 100 bp
* 99545 143263: contig of 43719 bp in length
* 143264 143363: gap of 100 bp
* 143364 192498: contig of 49135 bp in length.
FEATURES
    source
        1..192498
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /clone="RP24-301E12"
            /clone_lib="RPCI-24 Male Mouse BAC"
        1..1061
            /note="assembly_fragment"
        1162..5714
            /note="assembly_fragment"
        5815..11422
            /note="assembly_fragment"
        11523..17600
            /note="assembly_fragment"
        17701..30157
            /note="assembly_fragment"
        30258..38787
            /note="assembly_fragment"
        38888..99444
            /note="assembly_fragment"
        99545..143263
            /note="assembly_fragment"
        143364..192498
            /note="assembly_fragment
            clone end:77
            vector side:right"
BASE COUNT 53596 a 42471 c 43252 g 52379 t 800 others
ORIGIN
Query Match 86.2%; Score 106; DB 2; Length 192498;
Best Local Similarity 91.8%; Pred. No. 5,7e-26;
Matches 112; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

Qy 1 CAATTGAAAAAGTTTCTTCTAGTGGTTCGAAAGCCCAACACTGTGTCTTGTCCAGTGAG 60
 Db 90351 CACTGAAAAAGTTTCTTCTAGTGGTTCGAAAGCCCAACACTGTGTCTTGTCCAGTGAG 90410
 Qy 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCCAAGGT 120
 Db 90411 TTAGTTGTACAGACGACGATTCAGCACTAGCACTAGCGCTTGACAGAACCTCACAGACCCCAAGGA 90470
 Qy 121 AC 122
 Db 90471 AC 90472

RESULT 9
 AC132602 230015 bp DNA linear HTG 01-SEP-2002
 LOCUS Mus musculus chromosome UNK clone RP23-477C13, WORKING DRAFT
 DEFINITION SEQUENCE, 10 unordered pieces.
 ACCESSION AC132602
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 230015)
 AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE The sequence of Mus musculus clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 230015)
 AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: M.BA0477C13
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 225083 bases at least Q40
 Consensus quality: 225730 bases at least Q30
 Consensus quality: 226267 bases at least Q20
 Insert size: 202000; agarose-fp
 Insert size: 228405; sum-of-contigs
 Quality coverage: 13.11 in Q20 bases; agarose-fp
 Quality coverage: 9.64 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 710: contig of 710 bp in length
 * 711 810: gap of unknown length
 * 811 18316: contig of 17506 bp in length
 * 18317 18416: gap of unknown length
 * 18417 57380: contig of 38964 bp in length
 * 57381 57480: gap of unknown length
 * 57481 118173: contig of 60693 bp in length
 * 118174 118273: gap of unknown length

* 118274 201518: contig of 83245 bp in length
 * 201519 201618: gap of unknown length
 * 201619 202801: contig of 1183 bp in length
 * 202802 202901: gap of unknown length
 * 202902 204714: contig of 1813 bp in length
 * 204715 204814: gap of unknown length
 * 204815 208091: contig of 3277 bp in length
 * 208092 208191: gap of unknown length
 * 208192 213082: contig of 4891 bp in length
 * 213083 213183: gap of unknown length
 * 213183 230015: contig of 16833 bp in length.

FEATURES

source Location/Qualifiers
 1. 230015
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-477C13"
 1. 710
 /notes="assembly_name:Contig1"
 misc_feature 811..18316
 /note="assembly_name:Contig10"
 misc_feature 18417..57380
 /note="assembly_name:Contig11"
 misc_feature 57481..118173
 /note="assembly_name:Contig12"
 misc_feature 118274..201518
 /note="assembly_name:Contig13"
 misc_feature 201619..202801
 /note="assembly_name:Contig13"
 misc_feature 202902..204714
 /note="assembly_name:Contig16"
 misc_feature 204815..208091
 /note="assembly_name:Contig17"
 misc_feature 208192..213082
 /note="assembly_name:Contig18"
 misc_feature 213183..230015
 /note="assembly_name:Contig19"
 misc_feature 63025 a 50551 c 50278 g 65213 t 948 others
 BASE COUNT
 ORIGIN

Query Match 86.2%; Score 106; DB 2; Length 230015;
 Best Local Similarity 91.8%; Pred. No. 5.8e-26;

Matches 112; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACACTGTGTTCTTGCAGTGCAG 60
 |||||
 Db 109078 CAACGAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACACTGTATTTCTTGCAGTGCAG 109137

QY 61 TTAGTTGTACAGACGGCGTTAGCACTAGCGTTGCAGAACCTCAGACACCAAGGT 120
 |||||

Db 109138 TTAGTTGTACAGACGACGATTAGCACTAGCACTGTGCAGAACCTCAGACACCAAGGA 109197

QY 121 AC 122
 ||

Db 109198 AC 109199

RESULT 10

AC133022/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-50516, linear HTG 20-NOV-2002

***, 3 unordered pieces.

AC133022

AC133022.3 GI:25139057

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 228121)

AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Bisdwalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.I., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.B., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,B., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scheter,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sittler,C.D., Smajis,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczek,R., Wooden,H., Wright,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE

Direct Submission

JOURNAL

2 (bases 1 to 228121)

AUTHORS

Rat Genome Sequencing Consortium.

TITLE

Direct Submission

JOURNAL

3 (bases 1 to 228121)

AUTHORS

Rat Genome Sequencing Consortium.

TITLE

Direct Submission

JOURNAL

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 20, 2002 this sequence version replaced gi:23908439.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

COMMENT

On Sep 23, 2002 this sequence version replaced gi:21738449.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GRAL
 Center Clone name: CH230-257N6
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 196112 bases at least Q40
 Consensus quality: 198658 bases at least Q30
 Consensus quality: 200655 bases at least Q20
 Estimated insert size: 216180; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 120874: contig of 120874 bp in length
 * 120875 120974: gap of unknown length
 * 120975 173485: contig of 52511 bp in length
 * 173486 173585: gap of unknown length
 * 173586 192533: contig of 18948 bp in length
 * 192534 192633: gap of unknown length
 * 192634 193678: contig of 1045 bp in length
 * 193679 193778: gap of unknown length
 * 193779 194803: contig of 1025 bp in length
 * 194804 194903: gap of unknown length
 * 194904 197578: contig of 2675 bp in length
 * 197579 197678: gap of unknown length
 * 197679 200980: contig of 3302 bp in length
 * 200981 201080: gap of unknown length
 * 201081 214297: contig of 13217 bp in length
 * 214298 214397: gap of unknown length
 * 214398 217290: contig of 2893 bp in length
 * 217291 217390: gap of unknown length
 * 217391 226678: contig of 9288 bp in length
 * 226679 226778: gap of unknown length
 * 226779 236177: contig of 9299 bp in length
 * 236078 236177: gap of unknown length
 * 236178 239768: contig of 3591 bp in length.

FEATURES

source
 1. 239768
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-257N6"

BASE COUNT 62543 a 42292 c 41173 g 55976 t 37784 others
 ORIGIN

Query Match

Best Local Similarity

81.58; Score 100.2; DB 2; Length 239768;
 89.3%; Pred. No. 6.4e-24;

Matches 108; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 2 AATTGAAAAAGTTTCTTAGTGGTCGAAAGCCCAACACTGTGTTCTTGCAGTAGT 61
 Db 19799 AAGAAAAAAGTTTCTTAGTGGTTGAAGGCCCAACACTATATCTTGCAGTAGT 18858
 QY 62 TAGGTTGTACAGAACGGCGTTAGCACTAGCGTTTGACAGAACCTCAGACCCAAAGGTA 121
 Db 18859 TTGTTTACAGAAAGCGTTAGCACTAGCGTTTGACAGAACCTCAGACCCAAAGGAA 18918
 QY 122 C 122
 Db 18919 C 18919

RESULT 12.

AC130746

LOCUS

DEFINITION

Rattus norvegicus clone CH230-39M21, *** SEQUENCING IN PROGRESS

***, 3 unordered pieces.

AC130746

VERSION

AC130746.3 GI:30522353

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE

Rattus norvegicus

ORGANISM

Rattus norvegicus

REFERENCE

AUTHORS

Muzny, D., Maric, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Altschuler, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensunewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Nwankwelu, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Sma's, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 245468)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (14-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 245468)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23814855.

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGIN

Center clone name: CH230-39M21

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 236558 bases at least Q40

Consensus quality: 238736 bases at least Q30

Consensus quality: 240308 bases at least Q20

Estimated insert size: 249127; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 242398: contig of 242398 bp in length

* 242399 242498: gap of unknown length

* 242499 243549: contig of 1051 bp in length

* 243550 243649: gap of unknown length

* 243650 245468: contig of 1819 bp in length.

* Location/Qualifiers

1. 245468

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-39M21"

1. 1572

/notes="wgs end extension"

clone end:Sp6"

2074..3190

FEATURES

source

misc_feature

/note="wgs end extension"

clone end:Sp6"

2074..3190

misc_feature

/note="wgs end extension"

clone end:Sp6"

4356..15049

/note="clone boundary"

clone end:Sp6"

site:EcoRI

end sequence:BH286156"

238293..239134

/note="clone boundary"

clone end:T7

site:EcoRI

end sequence:BH286155"

BASE COUNT 70063 a 51006 c 48284 g 71652 t 4463 others

ORIGIN

Query Match 81.5%; Score 100.2; DB 2; Length 245468;

Best Local Similarity 89.3%; Pred. No. 6.4e-24;

Matches 108; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 AATTGAAAAAGTTTCTAGTGGTCTGAAAGCCCAACACTGTGTTCTTGCACGTAGT 61

Db 232776 AATGAAAAAGTTTCTAGTGGTCTGAAAGCCCAACACTATATATCTTGCACGTAGT 232835

Qy 62 TAGTTCTGACAGAACGCGTTAGCTAGCGCTTGACAGAACCTCAGACCCCAAGGTA 121

Db 232836 TTGTTACACAGATGCGTTAGCTAGCGTTTGACAGAACCTCAGACCCCAAGGTA 232895

Qy 122 C 122

Db 232896 C 232896

RESULT 13

AC123247

LOCUS

AC123247 Rattus norvegicus clone CH230-151E3, WORKING DRAFT SEQUENCE, 3

DEFINITION

unordered pieces.

ACCESSION

AC123247.4 GI:25138228

VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 240973)

AUTHORS

Muzny,D,Marle., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Georgiev,E., Geer,K., Gill,R., Garcia,A., Garner,T., Garza,M.,

Gebregeorgis,E., Geer,K., Gilly,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.I., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorensuhewa,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J.,

Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, M., Nundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Steidle, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, E., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 240973)
Worley, K.C.

Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 240973)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23097393.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTOE
Center clone name: CH230-151B3
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 230127 bases at least Q40
Consensus quality: 232322 bases at least Q30
Consensus quality: 233892 bases at least Q20
Estimated insert size: 236773; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will be preserved.
* 1 159537: contig of 159537 bp in length
* 159538 159637: gap of unknown length
* 159638 239730: contig of 79993 bp in length
* 239631 239730: gap of unknown length
* 239731 240973: contig of 1243 bp in length.

FEATURES
source
1. 240973
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-151B3"
misc_feature
1. 1047
/note="wgs contig"
misc_feature
3510..5688
/note="wgs contig"
BASE COUNT 57688 a 61245 c 61509 g 54580 t 5951 others
ORIGIN

Query Match 70.9%; Score 87.2; DB 2; Length 240973;
Best Local Similarity 87.7%; Pred. No. 2.4e-19;
Matches 107; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCAAGGGCCCAACACTGTGTCTTGTCCAGTGAG 60
|||||
Db 203927 CAATTGAAAAAGTTTGTCTAGGCGCTTAAAGGCCAACACTA--TTCTGCCAGTGAG 203984
|||||

QY 61 TTAGGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGACGCCCAAGGT 120
|||||
Db 203985 ATAGGTTATATAGATGGCGTCAGCACTAGCTGACAGAACCTCACAACCCAAAGGA 204044
|||||

QY 121 AC 122
Db 204045 AC 204046

RESULT 14
AC129824/c
LOCUS
DEFINITION
AC129824.3 GI:23265731
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 241048)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschuler, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, I., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Gebrageorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Lounsbeged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G., Olampunsaagoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rokey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanai, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

* This sequence will be replaced
 * By the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 241048: contig of 241048 bp in length.

Location/Qualifiers

1. 241048

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-88022"

1. 1324

/note="wgs_end_extension"

clone_end:Sp6"

4678..6863

/note="wgs_end_extension"

clone_end:Sp6"

complement(7042..7750)

/note="clone_boundary"

clone_end:Sp6"

site:EcoRI

end sequence:BH317537"

181147..182998

/note="wgs_contig"

234297..235118

/note="clone_boundary"

clone_end:T7

site:EcoRI

end sequence:BH317534"

238585..241048

/note="wgs_end_extension"

clone_end:T7"

BASE COUNT 51891 a 56832 c 56433 g 55074 t 20819 others

ORIGIN

Query Match

Best Local Similarity 70.9%; Score 87.2; DB 2; Length 241048;

Matches 107; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

Qy

1 CAATTGAAAAAGTTTCTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTCTAG 60

Db 156075 CAATTGAAAAAGTTTCTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTCTAG 156018

Qy

61 TTAGTTGTACAGAACGGGTAGCCTAGCGTTGACAGAACCTCAGACCCCAAGGT 120

Db 156017 ATAGTTTATAGAAATGGCGTCTAGCCTAGCAGTTGACAGAACCTCAGACCCCAAGGT 155958

Qy

121 AC 122

Db 155957 AC 155956

RESULT 15

AX163754/c

LOCUS AX163754

DEFINITION Sequence 18 from Patent WO0138579.

AX163754

ACCESSION AX163754

VERSION AX163754.1

KEYWORDS GI:14544860

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .352

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

Method of identifying toxic agents using nsaid-induced differential

gene expression in liver

Patent: WO 0138579-A 18 31-MAY-2001;

Curagen Corporation (US)

Location/Qualifiers

1. .352

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GSWG
 Center clone name: CH230-88022
 Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 215098 bases at least Q40
 Consensus quality: 217421 bases at least Q30
 Consensus quality: 218930 bases at least Q20
 Estimated insert size: 230359; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

BASE COUNT	86 a	77 c	72 g	117 t
ORIGIN				

Query Match 64.7%; Score 79.6; DB 6; Length 352;
Best Local Similarity 83.1%; Pred. No. 6.8e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

1 CAAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTCTTGTCCAGTGAG 60

D_b 216 CAATTAAATAATTCTGTTGTAGTGGCTGAAGGGTCCCACGCTGTATTCTCGCCAGTGAG 157

61 TTAGGTTGTACAGAACGGCGTTAGCCTAGCGC--TTGACAGAACCTCAGAGCCCAAG 118

Db
156 TTAAGTTGTACAGAACATCGTCAGCACTAGCACAGTTTACAGAACTTACAGACCCAAAG 97

Qy 119 GTAC 122

Db 96 GAAC 93

Search completed: November 23, 2003, 14:01:45
Job time : 468.915 secs


```
; OTHER INFORMATION: 20-841-149 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 45442
; OTHER INFORMATION: 20-842-115 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 77058
; OTHER INFORMATION: 20-853-415 : polymorphic base C or T
; NAME/KEY: primer_bind
; LOCATION: 929..949
; OTHER INFORMATION: 20-828.pu
; NAME/KEY: primer_bind
; LOCATION: 1357..1377
; OTHER INFORMATION: 20-828.rp complement
; NAME/KEY: primer_bind
; LOCATION: 12029..12050
; OTHER INFORMATION: 17-42.pu
; NAME/KEY: primer_bind
; LOCATION: 12581..12603
; OTHER INFORMATION: 17-42.rp complement
; NAME/KEY: primer_bind
; LOCATION: 14992..15012
; OTHER INFORMATION: 17-41.pu
; NAME/KEY: primer_bind
; LOCATION: 15460..15482
; OTHER INFORMATION: 17-41.rp complement
; NAME/KEY: primer_bind
; LOCATION: 42070..42090
; OTHER INFORMATION: 20-841.pu
; NAME/KEY: primer_bind
; LOCATION: 42572..42591
; OTHER INFORMATION: 20-841.rp complement
; NAME/KEY: primer_bind
; LOCATION: 45328..45347
; OTHER INFORMATION: 20-842.pu
; NAME/KEY: primer_bind
; LOCATION: 45863..45883
; OTHER INFORMATION: 20-842.rp complement
; NAME/KEY: primer_bind
; LOCATION: 76644..76664
; OTHER INFORMATION: 20-853.pu
; NAME/KEY: primer_bind
; LOCATION: 77166..77185
; OTHER INFORMATION: 20-853.rp complement
; NAME/KEY: primer_bind
; LOCATION: 1220..1238
; OTHER INFORMATION: 20-828-311.mis
; NAME/KEY: primer_bind
; LOCATION: 1240..1258
; OTHER INFORMATION: 20-828-311.mis complement
; NAME/KEY: primer_bind
; LOCATION: 12328..12346
; OTHER INFORMATION: 17-42-319.mis
; NAME/KEY: primer_bind
; LOCATION: 12348..12366
; OTHER INFORMATION: 17-42-319.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15222..15240
; OTHER INFORMATION: 17-41-250.mis
; NAME/KEY: primer_bind
; LOCATION: 15242..15260
; OTHER INFORMATION: 17-41-250.mis complement
; NAME/KEY: primer_bind
; LOCATION: 42199..42217
; OTHER INFORMATION: 20-841-149.mis
; NAME/KEY: primer_bind
; LOCATION: 42219..42237
; OTHER INFORMATION: 20-841-149.mis complement
; NAME/KEY: primer_bind
; LOCATION: 45423..45441
; OTHER INFORMATION: 20-842-115.mis
; NAME/KEY: primer_bind
; LOCATION: 45443..45461
; OTHER INFORMATION: 20-842-115.mis complement
```

```
; NAME/KEY: primer_bind
; LOCATION: 77039..77057
; OTHER INFORMATION: 20-853-415.mis
; NAME/KEY: primer_bind
; LOCATION: 77059..77077
; OTHER INFORMATION: 20-853-415.mis complement
; NAME/KEY: misc_binding
; LOCATION: 1227..1251
; OTHER INFORMATION: 20-828-311.probe
; NAME/KEY: misc_binding
; LOCATION: 12335..12359
; OTHER INFORMATION: 17-42-319.probe
; NAME/KEY: misc_binding
; LOCATION: 15229..15253
; OTHER INFORMATION: 17-41-250.probe
; NAME/KEY: misc_binding
; LOCATION: 42206..42230
; OTHER INFORMATION: 20-841-149.probe
; NAME/KEY: misc_binding
; LOCATION: 45430..45454
; OTHER INFORMATION: 20-842-115.probe
; NAME/KEY: misc_binding
; LOCATION: 77046..77070
; OTHER INFORMATION: 20-853-415.probe
; US-09-750-580-1
```

```
Query Match 21.6%; Score 26.6; DB 4; Length 81001;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 38; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
Qy 6 GAAAAAGTTGTTCTAGTGTGGAAGGCCAACACACTGTCTTCTGCAGTGAGTT 62
Db 77006 GAAAGGACTTGTGTTGGGTGGCAGGAGTAACACTGTCTTTAGAAAGAYCGTT 77062
```

RESULT 5

```
US-09-252-991A-778
; Sequence 778, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 778
; LENGTH: 786
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-778
```

```
Query Match 21.5%; Score 26.4; DB 4; Length 786;
Best Local Similarity 57.1%; Pred. No. 2.1;
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
Qy 15 TTGTTCTAGTGTGCGAAAGGCCAACACACTGTCTTCTGCAGTGAGTTAGTTACAGA 74
Db 670 TTGATCAGGTGTTGGAGGTGCCATCTTCTCTTCTCTTCGCTAGGGGTGACGAA 729
Qy 75 ACGGCGTTAGCACTAGCGTTGAC 98
Db 730 AGGCGCGTGGCCCTTTCCTTTAC 753
```

RESULT 6

```
US-09-252-991A-638/c
; Sequence 638, Application US/09252991A
```

Db 843 NGNAAATGYGCGNTG 857

RESULT 8

US-09-706-968-6

Sequence 6, Application US/09706968

Patent No. 6528050

GENERAL INFORMATION:

APPLICANT: Gao, Zeren

APPLICANT: Hart, Charles E.

APPLICANT: Piddington, Christopher S.

APPLICANT: Sheppard, Paul O.

APPLICANT: Shoemaker, Kimberly E.

APPLICANT: Gilbertson, Debra G.

APPLICANT: West, James W.

TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3

FILE REFERENCE: 98-60C1

CURRENT APPLICATION NUMBER: US/09/706,968

CURRENT FILING DATE: 2000-11-06

PRIOR APPLICATION NUMBER: US/09/541,752

PRIOR FILING DATE: 2000-03-31

NUMBER OF SEQ ID NOS: 50

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 6

LENGTH: 1035

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: degenerate sequence derived from SEQ ID NOS: 1 and 2

NAME/KEY: misc feature

LOCATION: (1)..(1035)

OTHER INFORMATION: n = A,T,C or G

US-09-706-968-6

Query Match 21.5%; Score 26.4; DB 4; Length 900;

Best Local Similarity 57.1%; Pred. No. 2.2;

Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 15 TTGTTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTAGTTGTACAGA 74

Db 159 TTGATCAGGTGGTGGAGGGGCGCCATCTGTTCTGTCTCTCTCTCGCTAGGGGTGACGAA 100

QY 75 ACGCGCTTAGCACTAGCGCTTGAC 98

Db 99 AGGCGGTGGCCCTTTCCCTTTAC 76

RESULT 7

US-09-457-066-6

Sequence 6, Application US/09457066

Patent No. 6432673

GENERAL INFORMATION:

APPLICANT: Gao, Zeren

APPLICANT: Hart, Charles E.

APPLICANT: Piddington, Christopher S.

APPLICANT: Sheppard, Paul O.

APPLICANT: Shoemaker, Kimberly E.

APPLICANT: Gilbertson, Debra G.

APPLICANT: West, James W.

TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3

FILE REFERENCE: 98-60

CURRENT APPLICATION NUMBER: US/09/457,066

CURRENT FILING DATE: 1999-12-07

NUMBER OF SEQ ID NOS: 50

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 6

LENGTH: 1035

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: degenerate sequence derived from SEQ ID NOS: 1 and 2

NAME/KEY: misc feature

LOCATION: (1)..(1035)

OTHER INFORMATION: n = A,T,C or G

US-09-457-066-6

Query Match 21.5%; Score 26.4; DB 4; Length 1035;

Best Local Similarity 40.0%; Pred. No. 2.3;

Matches 30; Conservative 13; Mismatches 32; Indels 0; Gaps 0;

QY 22 AGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTAGTTGTACAGACGCGT 81

Db 783 RGARYTNAARMGNACNGAYACNATHITYTGCCNGGNTGYTYNTGTNAARMGNTGYGG 842

QY 82 TAGCACTAGCGCTTG 96

Db 843 NGNAAATGYGCGNTG 857

Query Match 21.5%; Score 26.4; DB 4; Length 1035;

Best Local Similarity 40.0%; Pred. No. 2.3;

Matches 30; Conservative 13; Mismatches 32; Indels 0; Gaps 0;

QY 22 AGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTAGTTGTACAGACGCGT 81

Db 783 RGARYTNAARMGNACNGAYACNATHITYTGCCNGGNTGYTYNTGTNAARMGNTGYGG 842

QY 82 TAGCACTAGCGCTTG 96

Db 843 NGNAAATGYGCGNTG 857

RESULT 9

US-09-328-352-2269

Sequence 2269, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 2269

LENGTH: 1572

TYPE: DNA

ORGANISM: Acinetobacter baumannii

US-09-328-352-2269

Query Match 21.5%; Score 26.4; DB 4; Length 1572;

Best Local Similarity 52.8%; Pred. No. 2.8;

Matches 57; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 14 TTGTTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTAGTTGTACAG 73

Db 1342 TTGTCACAGATTGGTTTACTTCCAACACTGACTTCTTAAAGACAGTGAGGTTGAGTTA 1401


```
; 74 AACGGCTTAGCTAGCTAGCGTTGACAGAACCTCAGACACCCCAAGGTA 121
; 1402 ACCAACCGTGGCGAGATCATTTGTTAATGACCGCAACGAAACCAATGTA 1449

RESULT 10
; Sequence 1, Application US/09214808A
; Patent No. 6475793
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; Patent No. 6475793
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0068
; CURRENT APPLICATION NUMBER: US/09/214,808A
; CURRENT FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: PCT/IB97/00950
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
; US-09-214-808-1

Query Match 21.5%; Score 26.4; DB 4; Length 536165;
Best Local Similarity 54.0%; Pred. No. 27; Gaps 0;
Matches 54; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 17 GTTCTAGTGTGAAAGGCCCAACTGTGTCTTGCCAGTGAGTTAGTGTGACAGAAC 76
Db 392324 GATGGAGTGTTAGTAGACCCATCACGTTGATTTCCAGAGAGTATCTCGCTCAAG 392383

QY 77 GCGTTAGCTAGCTAGCGTTGACAGAACCTCAGACACCCAA 116
Db 392384 AGCGATTCAACCGCGCACTAGGCGCAACTCAGAGAGACAA 392423

RESULT 11
; US-08-462-080B-1
; Sequence 1, Application US/08462080B
; Patent No. 5997913
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Shoemaker, Sharon
; APPLICANT: Barnett, Christopher C.
; TITLE OF INVENTION: Saccharification of Cellulose by Cloning and
; TITLE OF INVENTION: Amplification of the Beta-glucosidase Gene of Trichoderma Reesei
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: Ca
; COUNTRY: U.S.A.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,080B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/248,586
```

```
; FILING DATE: 24-MAY-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,028
; FILING DATE: 10-DEC-1991
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/625,140
; FILING DATE: 10-DEC-1990
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,656
; REFERENCE/DOCKET NUMBER: GC78D3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7555
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoderma reesei
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(311...375, 446...2205, 2270...2679)
; FEATURE:
; NAME/KEY: intron
; LOCATION: 376...445
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2206...2269
; US-08-462-080B-1

Query Match 20.8%; Score 25.6; DB 2; Length 3033;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 49; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 12 AGTTTGTCTAGTGTGCGAAAGGCCCAACTGTGTCTTGCAGTGAGTTAGTTGTAC 71
Db 260 AGTTCTCTTACCAAGTCTTGACCAAGACCATCTGTGTAGCCCAATCAGAAATGCGTTAC 319

QY 72 AGAACGCGGTAGCACTAGCGCTTGACA 99
Db 320 CGAACAGACGCTCGCTGCGCACTTGCCA 347

RESULT 12
; US-08-462-090-1
; Sequence 1, Application US/08462090
; Patent No. 6022725
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Barnett, Christopher C.
; APPLICANT: Shoemaker, Sharon
; TITLE OF INVENTION: Saccharification of Cellulose by Cloning
; TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Building, 699 Prince St.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22313-1404
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,090
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/625,140
FILING DATE: 10-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Dillahunty, T. Gene
REGISTRATION NUMBER: 25,423
REFERENCE/DOCKET NUMBER: 010055-056
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
FEATURE:
NAME/KEY: CDS
LOCATION: join(311..375, 446..2205, 2270..2679)
FEATURE:
NAME/KEY: intron
LOCATION: 376..445
FEATURE:
NAME/KEY: intron
LOCATION: 2206..2269
US-08-462-090-1

Query Match 20.8%; Score 25.6; DB 3; Length 3033;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 49; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 12 AGTTGTTCTTAGTGGTCGAAAGGCCCAACACTGTGTTGCGCAGTGAGTTAGGTTGTC 71
Db 260 AGTTCTTCTTACAGTCTTGACCAAGACCATTCTGTGAGCCCAATCAGAAATGCGTTAC 319
QY 72 AGAAGCGCGTTAGCACTAGCGCTTGACA 99
Db 320 CGAAGACAGAGCTGCGCTGGCACTTGCCA 347

RESULT 13
US-08-463-461-1
Sequence 1, Application US/08463461
Patent No. 6103464
GENERAL INFORMATION:
APPLICANT: Fowler, Timothy
APPLICANT: Barnett, Christopher C.
APPLICANT: Shoemaker, Sharon
TITLE OF INVENTION: Saccharification of Cellulose by Cloning
PRIOR APPLICATION: and Amplification of the Beta-glucosidase Gene of
TITLE OF INVENTION: Trichoderma Reesei
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,461
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC78D4
TELEPHONE: 415-846-7555
TELEFAX: 415-845-6504
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
FEATURE:
NAME/KEY: CDS
LOCATION: join(311..375, 446..2205, 2270..2679)
FEATURE:
NAME/KEY: intron
LOCATION: 376..445
FEATURE:
NAME/KEY: intron
LOCATION: 2206..2269
US-08-463-461-1

Query Match 20.8%; Score 25.6; DB 3; Length 3033;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 49; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 12 AGTTGTTCTTAGTGGTCGAAAGGCCCAACACTGTGTTGCGCAGTGAGTTAGGTTGTC 71
Db 260 AGTTCTTCTTACAGTCTTGACCAAGACCATTCTGTGAGCCCAATCAGAAATGCGTTAC 319
QY 72 AGAAGCGCGTTAGCACTAGCGCTTGACA 99
Db 320 CGAAGACAGAGCTGCGCTGGCACTTGCCA 347

RESULT 14
US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature

LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1667854)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 20.7%; Score 25.4; DB 4; Length 1664976;
Best Local Similarity 58.7%; Pred. No. 84;
Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 10 AAAGTTTGTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTGTAGTTGTACAGAACGGCGT 81
Db 1601936 AATATTTATGCTGTGTGTAAGGCCCAACACTGTGTTCTTGCCAGTGTAGTTGTCT 59
QY 70 ACAGAACGGCGCTTAG 84
Db 1601876 CTTTAACTGCGAGAG 1601862

RESULT 15
US-08-614-770A-1/c
Sequence 1, Application US/08614770A
Patent No. 5773267
GENERAL INFORMATION:
APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,770A
FILING DATE: MARCH 7, 1996
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/402
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49272
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OLIGONUCLEOTIDE
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIOPHAGE
INDIVIDUAL ISOLATE: D29
US-08-614-770A-1

Query Match 20.5%; Score 25.2; DB 1; Length 49272;
Best Local Similarity 55.8%; Pred. No. 30;
Matches 48; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 22 AGTGGTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGTAGTTGTACAGAACGGCGT 81
Db 47920 AGTGGTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGTAGTTGTACAGAACGGCGT 81
QY 82 TAGCACTAGCGCTTGACAGAACCTCA 107
Db 47860 CATCAGCATCACTGGCGACACCGTCA 47835

Search completed: November 23, 2003, 15:53:39
Job time : 19.3791 secs

C 1	79.6	64.7	1232	10	US-09-680-107-1632	Sequence 1632, Ap
C 2	79.6	64.7	1644	14	US-10-153-668-48	Sequence 48, Appl
C 3	79.6	64.7	1646	14	US-10-153-668-46	Sequence 46, Appl
C 4	79.6	64.7	2051	9	US-09-925-302-255	Sequence 255, Appl
C 5	79.6	64.7	2941	14	US-10-198-846-9874	Sequence 9874, Ap
C 6	57	46.3	353	10	US-09-960-352-11432	Sequence 11432, Ap
C 7	38.2	31.1	3740	11	US-09-764-891-9984	Sequence 9984, Ap
C 8	33.4	27.2	2313	14	US-10-198-846-12319	Sequence 12319, A
C 9	31.4	25.5	477	14	US-10-066-543-3022	Sequence 3022, Ap
C 10	31.4	25.5	549	14	US-10-066-543-2878	Sequence 2878, Ap
C 11	31.4	25.5	567	14	US-10-066-543-2964	Sequence 2964, Ap
C 12	31.4	25.5	602	13	US-09-878-178-1654	Sequence 1654, Ap
C 13	31.4	25.5	602	13	US-10-046-935-1554	Sequence 1554, Ap
C 14	31.4	25.5	602	14	US-10-146-503-1854	Sequence 1854, Ap
C 15	33.4	25.5	605	11	US-09-071-161-483	Sequence 483, App
C 16	29.4	23.9	509	12	US-10-027-632-323187	Sequence 323187,

QY 119 GTAC 122
Db 1015 GAAC 1012

RESULT 2

US-10-153-668-48/c
; Sequence 48, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
US-10-153-668-48

Query Match 64.7%; Score 79.6; DB 14; Length 1644;
Best Local Similarity 83.1%; Pred. No. 3.1e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 60
Db 1548 CAATTTAAAAAATTCCTGTAGTGGCTGAAGGGTCCCGCTGTATTCTCGCCAGTGAG 1489
QY 61 TTAGTTGTACAGAACGGCTTAGCACTAGCGC--TTGACAGAACCTTCAGACCCCAAAG 118
Db 1488 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACCTTCAGACCCCAAAG 1429
QY 119 GTAC 122
Db 1428 GAAC 1425

RESULT 3

US-10-153-668-46/c
; Sequence 46, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25

; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(273)
US-10-153-668-46
Query Match 64.7%; Score 79.6; DB 14; Length 1646;
Best Local Similarity 83.1%; Pred. No. 3.1e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 60
Db 1550 CAATTTAAAAAATTCCTGTAGTGGCTGAAGGGTCCCGCTGTATTCTCGCCAGTGAG 1491
QY 61 TTAGTTGTACAGAACGGCTTAGCACTAGCGC--TTGACAGAACCTTCAGACCCCAAAG 118
Db 1490 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACCTTCAGACCCCAAAG 1431
QY 119 GTAC 122
Db 1430 GAAC 1427

RESULT 4

US-09-925-302-255/c
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P4104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)

```
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-255

Query Match      64.7%; Score 79.6; DB 9; Length 2051;
Best Local Similarity 83.1%; Pred. No. 3.4e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1996 CAATTGAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 1837

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAAACCTCACAGACCCAAAG 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1836 TTAGTTGTACAGAACATCGTCAGCACTAGCACTAGCACTTACAGAACTTACAGACCCAAAG 1777

QY 119 GTAC 122
Db ||||
1776 GAAC 1773

RESULT 5
US-10-198-846-9874/c
; Sequence 9874, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 9874
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-9874

Query Match      64.7%; Score 79.6; DB 14; Length 2941;
Best Local Similarity 83.1%; Pred. No. 3.9e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2336 CAATTGAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 2277

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAAACCTCACAGACCCAAAG 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2276 TTAGTTGTACAGAACATCGTCAGCACTAGCACTAGCACTTACAGAACTTACAGACCCAAAG 2217

QY 119 GTAC 122
Db ||||
2216 GAAC 2213

RESULT 6
US-09-960-352-11432
; Sequence 11432, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
```

```
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11432
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 49-LIB3058-009-Q1-K1-E2
US-09-960-352-11432

Query Match      46.3%; Score 57; DB 10; Length 353;
Best Local Similarity 82.8%; Pred. No. 9.9e-12;
Matches 77; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 CAATTGAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 CAATTGAAAAGTTTCTTCTAGTGGTTGAAAGTCCCAAGCTCGTATCTTGCAGTG-G 213

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC 93
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 GTAAGTTGTACAGAACTTCGTTAGCAGGAC 246

RESULT 7
US-09-764-891-9984/c
; Sequence 9984, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9984
; LENGTH: 3740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9984

Query Match      31.1%; Score 38.2; DB 11; Length 3740;
Best Local Similarity 78.0%; Pred. No. 0.00048;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 8 AAAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAGTTAGT 66
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2725 AAAAAATTTGTTCTAGTTGTTGAAAGGCCCAAGCTTATATCTTGCAGTGCTTAAGT 2667

RESULT 8
US-10-198-846-12919/c
; Sequence 12919, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 12919
; LENGTH: 2313
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-12919

Query Match      27.2%; Score 33.4; DB 14; Length 2313;
Best Local Similarity 86.0%; Pred. No. 0.029;
Matches 49; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 68 GTACAGACGGGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2296 GTACAGACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 2240

RESULT 9
US-10-066-543-3022
; Sequence 3022, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3022
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-3022

Query Match      25.5%; Score 31.4; DB 14; Length 477;
Best Local Similarity 85.5%; Pred. No. 0.093;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 70 ACAGAACGGGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 55

RESULT 10
US-10-066-543-2878
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2878

Query Match      25.5%; Score 31.4; DB 14; Length 567;
Best Local Similarity 85.5%; Pred. No. 0.099;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 70 ACAGAACGGGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 55

RESULT 11
US-10-066-543-2964
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2964

Query Match      25.5%; Score 31.4; DB 14; Length 567;
Best Local Similarity 85.5%; Pred. No. 0.098;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 70 ACAGAACGGGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 55

RESULT 12
US-09-878-178-1654
; Sequence 1654, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
```



```
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(602)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1654
```

```
Query Match 25.5%; Score 31.4; DB 10; Length 602;
Best Local Similarity 85.5%; Pred. No. 0.1;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;
```

```
QY 70 ACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACCTTCACAGACCCCAAAGGTAC 122
Db 1 ACAGACATCGTCAGCACTAGCAGTTTACAGAACCTTCACAGACCCCAAAGGAAC 55
```

RESULT 13

```
US-10-046-935-1654
; Sequence 1654, Application US/10046935
; Publication No. US2002015601A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 569, 578
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-1654
```

```
Query Match 25.5%; Score 31.4; DB 13; Length 602;
Best Local Similarity 85.5%; Pred. No. 0.1;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;
```

```
QY 70 ACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACCTTCACAGACCCCAAAGGTAC 122
Db 1 ACAGACATCGTCAGCACTAGCAGTTTACAGAACCTTCACAGACCCCAAAGGAAC 55
```

RESULT 14

```
US-10-146-502-1654
; Sequence 1654, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
```

```
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 569, 578
; OTHER INFORMATION: n = A,T,C or G
US-10-146-502-1654
```

```
Query Match 25.5%; Score 31.4; DB 14; Length 602;
Best Local Similarity 85.5%; Pred. No. 0.1;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;
```

```
QY 70 ACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACCTTCACAGACCCCAAAGGTAC 122
Db 1 ACAGACATCGTCAGCACTAGCAGTTTACAGAACCTTCACAGACCCCAAAGGAAC 55
```

RESULT 15

```
US-09-871-161-483
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/871,161
; CURRENT FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 09/328,111
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/117,393
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: 60/098,639
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483
```

```
Query Match 25.5%; Score 31.4; DB 11; Length 605;
Best Local Similarity 85.5%; Pred. No. 0.1;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;
```

```
QY 70 ACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACCTTCACAGACCCCAAAGGTAC 122
Db 1 ACAGACATCGTCAGCACTAGCAGTTTACAGAACCTTCACAGACCCCAAAGGAAC 55
```

Search completed: November 23, 2003, 19:05:39
Job time : 47.5904 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 23, 2003, 09:18:57 : Search time 3810.98 Seconds
(without alignments)
10917.162 Million cell updates/sec

Title: US-09-717-321A-15
Perfect score: 1017
Sequence: 1 cccctattctgtcagatt.....cctttgggtctgtgaggttc 1017

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 577422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1017	100.0	1017	6	AX163751	AX163751 Sequence
c	1017	100.0	1266	6	AX163738	AX163738 Sequence
3	1013.8	99.7	217700	2	AC106124	AC106124 Rattus no
4	679.2	66.8	2281	10	BC051053	BC051053 Mus muscu
5	677.6	66.6	2319	10	BC003828	BC003828 Mus muscu
6	677.6	66.6	269081	2	AC068493	AC068493 Mus muscu
7	403.2	39.6	185245	2	AC102775	AC102775 Mus muscu
8	403.2	39.6	202342	10	AL583884	AL583884 Mouse DNA
9	337.2	33.2	179685	10	AC126556	AC126556 Mus muscu
10	290.6	28.6	1022	6	AX163752	AX163752 Sequence
11	290.6	28.6	2302	9	BC050687	BC050687 Homo sapi
12	290.6	28.6	2315	9	AK054993	AK054993 Homo sapi
13	290.6	28.6	28567	9	HSAL32695	HSAL32695 Homo sapi
14	290.6	28.6	212827	9	AC009412	AC009412 Homo sapi
15	285.2	28.0	1232	6	AX408985	AX408985 Sequence
16	285.2	28.0	1232	9	HMP02879	D52774 Homo sapien
c	173	19.1	202565	9	AL354696	AL354696 Human DNA
c	193.2	19.0	605	6	AX341407	AX341407 Sequence
c	190.2	18.7	192498	2	AC105979	AC105979 Mus muscu
c	190.2	18.7	230015	2	AC132602	AC132602 Mus muscu
21	173	17.0	137625	9	AC104663	AC104663 Homo sapi
c	169.8	16.7	5544	9	AF542527	AF542527 Homo sapi
c	116.2	11.4	137625	9	AC104663	AC104663 Homo sapi
c	113.8	11.2	87616	2	AC139405	AC139405 Homo sapi
25	113	11.1	348	6	AX163753	AX163753 Sequence
26	112.4	11.1	64781	2	CNS01DS5	AL121655 BAC seque
c	105.8	10.4	101584	9	AC012364	AC012364 Homo sapi
c	105.8	10.4	155943	9	AC012364	AC012364 Homo sapi
29	103.4	10.2	5544	9	AF542527	AF542527 Homo sapi
30	103.2	10.1	110816	9	AC002404	AC002404 Human Chr
c	98.4	9.7	170839	2	AC133467	AC133467 Mus muscu
c	85.6	8.4	495	6	AX341256	AX341256 Sequence
c	84.2	8.3	180303	9	AL672045	AL672045 Human DNA
c	84.2	8.3	201012	2	AC021189	AC021189 Homo sapi
35	76.2	7.5	228121	2	AC133022	AC133022 Rattus no
c	76.2	7.5	239768	2	AC112582	AC112582 Rattus no
c	76.2	7.5	245468	2	AC130746	AC130746 Rattus no
38	67	6.6	594	6	AX385363	AX385363 Sequence
c	62.8	6.2	194371	2	AC135453	AC135453 Rattus no
40	61.4	6.0	378	6	BD030694	BD030694 Sequence
41	56.8	5.6	198	11	G31747	G31747 swSS2749 Er
42	54.2	5.3	7218	6	I66494	I66494 Sequence 14
43	52.4	5.2	162	6	AX397334	AX397334 Sequence
c	51.4	5.1	90354	9	AP001124	AP001124 Homo sapi
45	51.4	5.1	178089	2	AC010929	AC010929 Homo sapi

ALIGNMENTS

RESULT 1
AX163751
LOCUS AX163751 1017 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 15 from Patent WO0138579.
ACCESSION AX163751
VERSION AX163751.1 GI:14544857
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Gould-Rothberg,B.E., Dipippo,V.A., Ramseh,T.M. and Gerwein,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential

gene expression in liver
Patent: WO 0138579-A 15 31-MAY-2001;
Curagen Corporation (US)
Location/Qualifiers
1. .1017
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
BASE COUNT 245 a 245 c 216 g 311 t
ORIGIN

Query Match 100.0%; Score 1017; DB 6; Length 1017;
Best Local Similarity 100.0%; Pred. No. 1.2e-291;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATTCCTGCTCAGATTAGAAATGCGGAGCTTGTGAGAGCTTTGCTTAAAGAGACTGCAGCTTC 60
Db 1 CCCCTATTCCTGCTCAGATTAGAAATGCGGAGCTTGTGAGAGCTTTGCTTAAAGAGACTGCAGCTTC 60

QY 61 CTGAGAACAGCTTAAGCACTAAGCTGTGAGAGACTTTGCTTAAAGAGACTGCAGCTTC 120
Db 61 CTGAGAACAGCTTAAGCACTAAGCTGTGAGAGACTTTGCTTAAAGAGACTGCAGCTTC 120

QY 121 TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCGAGACCGCTGTGACACAGCAGCGCTCC 180
Db 121 TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCGAGACCGCTGTGACACAGCAGCGCTCC 180

QY 181 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 240
Db 181 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 240

QY 241 TTGTGACTGTACCTCAGGATGGGTGTAAAGCTCTGCTTCTTATTTTATGAGTGTCT 300
Db 241 TTGTGACTGTACCTCAGGATGGGTGTAAAGCTCTGCTTCTTATTTTATGAGTGTCT 300

QY 301 CTAAATATACAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTGGCTTCCTGTGTTC 360
Db 301 CTAAATATACAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTGGCTTCCTGTGTTC 360

QY 361 CTCTAAGCAAGTATTCCTGCTCAGCTGAGTGTGCTGGGTGAGTGTGTAACACGA 420
Db 361 CTCTAAGCAAGTATTCCTGCTCAGCTGAGTGTGCTGGGTGAGTGTGTAACACGA 420

QY 421 CGTCATCAAGGAGACAGACAGTATTTGACTAATATGAAGTAGAGATTAATTTACACTA 480
Db 421 CGTCATCAAGGAGACAGACAGTATTTGACTAATATGAAGTAGAGATTAATTTACACTA 480

QY 481 CATTGTACATGGAGTAACTCAACTGAATAAAAGTGTCACGGGTAAAGCTTTTAAACGGTT 540
Db 481 CATTGTACATGGAGTAACTCAACTGAATAAAAGTGTCACGGGTAAAGCTTTTAAACGGTT 540

QY 541 AATTTCTGTAACAGTAGATGACAAATGCCGATCTTATCAGTGTCTCTTTGAGCCCC 600
Db 541 AATTTCTGTAACAGTAGATGACAAATGCCGATCTTATCAGTGTCTCTTTGAGCCCC 600

QY 601 CTTTCCCTCGTCTCCCTCCAGATGGGGGTGAGTCCATATTTAAACTGGCCATCC 660
Db 601 CTTTCCCTCGTCTCCCTCCAGATGGGGGTGAGTCCATATTTAAACTGGCCATCC 660

QY 661 TCACAGTTGCTAATCTAGCAAGTCTTTCTTTAGGACCCCTCTTTTAAAGCAATATG 720
Db 661 TCACAGTTGCTAATCTAGCAAGTCTTTCTTTAGGACCCCTCTTTTAAAGCAATATG 720

QY 721 TCTGACCTGTACTATAGATCTTTCTGATATGCAATTCGGAGATTTTTTGGTAGATG 780
Db 721 TCTGACCTGTACTATAGATCTTTCTGATATGCAATTCGGAGATTTTTTGGTAGATG 780

QY 781 AGAAGTCGCTTCTGTTTTCACCTTCCTTTACTCAGCTGACTAGTCTTCCCTTCGTTTT 840
Db 781 AGAAGTCGCTTCTGTTTTCACCTTCCTTTACTCAGCTGACTAGTCTTCCCTTCGTTTT 840

QY 841 CTAGTAACCTGGGTAGAAATCAGTCTCGGCTTTTACAGTTTTTTAAACTATTTTAGAT 900

gene expression in liver
Patent: WO 0138579-A 15 31-MAY-2001;
Curagen Corporation (US)
Location/Qualifiers
1. .1017
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
BASE COUNT 245 a 245 c 216 g 311 t
ORIGIN

Query Match 100.0%; Score 1017; DB 6; Length 1266;
Best Local Similarity 100.0%; Pred. No. 1.3e-291;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATTCCTGCTCAGATTAGAAATGCGGAGCTTGTGAGAGCTTTGCTTAAAGAGACTGCAGCTTC 60
Db 1261 CCCTATTCCTGCTCAGATTAGAAATGCGGAGCTTGTGAGAGCTTTGCTTAAAGAGACTGCAGCTTC 1202

QY 61 CTGAGAACAGCTTAAGCACTAAGCTGTGAGAGACTTTGCTTAAAGAGACTGCAGCTTC 120
Db 1201 CTGAGAACAGCTTAAGCACTAAGCTGTGAGAGACTTTGCTTAAAGAGACTGCAGCTTC 1142

QY 121 TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCGAGACCGCTGTGACACAGCAGCGCTCC 180
Db 1141 TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCGAGACCGCTGTGACACAGCAGCGCTCC 1082

QY 181 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 240
Db 1081 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 1022

QY 241 TTGTGACTGTACCTCAGGATGGGTGTAAAGCTCTGCTTCTTATTTTATGAGTGTCT 300
Db 1021 TTGTGACTGTACCTCAGGATGGGTGTAAAGCTCTGCTTCTTATTTTATGAGTGTCT 962

QY 301 CTAAATATACAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTGGCTTCCTGTGTTC 360
Db 961 CTAAATATACAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTGGCTTCCTGTGTTC 902

QY 361 CTCTAAGCAAGTATTCCTGCTCAGCTGAGTGTGCTGGGTGAGTGTGTAACACGA 420
Db 901 CTCTAAGCAAGTATTCCTGCTCAGCTGAGTGTGCTGGGTGAGTGTGTAACACGA 842

QY 421 CGTCATCAAGGAGACAGACAGTATTTGACTAATATGAAGTAGAGATTAATTTACACTA 480
Db 841 CGTCATCAAGGAGACAGACAGTATTTGACTAATATGAAGTAGAGATTAATTTACACTA 782

Db 841 CTAGTAACCTGGGTAGAAATCAGCTCTCGGCTTTTACAGTTTTTTAAACTATTTTAGAT 900

QY 901 ATTCTGAACATCACTGTCTTCCAGAGTACCAAGACTGTCACTGATTGATCCGCCCC 960
Db 901 ATTCTGAACATCACTGTCTTCCAGAGTACCAAGACTGTCACTGATTGATCCGCCCC 960

QY 961 CTCTAGACCTCAACCCACGGGACACATGCTTCGGTACCTTTGGGTCTGTGAGGTTTC 1017
Db 961 CTCTAGACCTCAACCCACGGGACACATGCTTCGGTACCTTTGGGTCTGTGAGGTTTC 1017

RESULT 2
AX163738/c
LOCUS AX163738 1266 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 2 from Patent WO0138579.
ACCESSION AX163738
VERSION AX163738.1 GI:14544844
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Gould-Rothberg,B.E., Dippippo,V.A., Ramseh,T.M. and Gerwein,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
JOURNAL gene expression in liver
Patent: WO 0138579-A 2 31-MAY-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
1. .1266
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
BASE COUNT 385 a 258 c 285 g 338 t
ORIGIN

* consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 109797: contig of 109797 bp in length
 * 109798: gap of unknown length
 * 109799: gap of unknown length
 * 166934: contig of 57037 bp in length
 * 166935: gap of unknown length
 * 167034: gap of unknown length
 * 167035: contig of 9350 bp in length
 * 176384: contig of 9350 bp in length
 * 176385: gap of unknown length
 * 176484: gap of unknown length
 * 206328: contig of 29843 bp in length
 * 206428: gap of unknown length
 * 207462: contig of 1034 bp in length
 * 207562: gap of unknown length
 * 208607: contig of 1046 bp in length
 * 208608: gap of unknown length
 * 208708: gap of unknown length
 * 209908: contig of 1201 bp in length
 * 210008: gap of unknown length
 * 210009: contig of 1575 bp in length
 * 211583: gap of unknown length
 * 211584: contig of 1966 bp in length
 * 213649: contig of 1966 bp in length
 * 213650: gap of unknown length
 * 213750: contig of 1366 bp in length
 * 215116: gap of unknown length
 * 215216: contig of 2485 bp in length.

FEATURES

Source

1. 217700
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-119E11"

BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others
 ORIGIN

Query Match 99.7%; Score 1013.8; DB 2; Length 217700;
 Best Local Similarity 99.8%; Pred. No. 1.7e-290;
 Matches 1015; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCTATTCTGCTCAGATTAGAAATGCCAAATACCTTGTAATAGTTCGGTTGTG 60
 DB CCCCTATTCTGCTCAGATTAGAAATGCCAAATACCTTGTAATAGTTCGGTTGTG 63506

QY 61 CTGAGAACGTAAGCACTAAGCTGTGAGAGACTTTGCTCTTAAAGAACTGCAGCTTC 120
 DB CTGAGAACGTAAGCACTAAGCTGTGAGAGACTTTGCTCTTAAAGAACTGCAGCTTC 63566

QY 121 TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCAGACCGTGTGACACAGCGCTCC 180
 DB TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCAGACCGTGTGACACAGCGCTCC 63626

QY 181 TTAATGACACGCTGCCATGTAAAGCACTGTAATCTATAGCCCATGCTCATTAACGTAAC 240
 DB TTAATGACACGCTGCCATGTAAAGCACTGTAATCTATAGCCCATGCTCATTAACGTAAC 63686

QY 241 TTGTGACTGACCTACGATGGGTGTAAAGCTCTGCTCTTTGATTTCATGAGTTCT 300
 DB TTGTGACTGACCTACGATGGGTGTAAAGCTCTGCTCTTTGATTTCATGAGTTCT 63746

QY 301 CTAATAATACAGCTGACCGCTCTCAGCGCTTTGAACAGAACTCTGGCTCTGTTGTC 360
 DB CTAATAATACAGCTGACCGCTCTCAGCGCTTTGAACAGAACTCTGGCTCTGTTGTC 63806

QY 361 CTCTAACGAAGTATTCTGTTCTTAGTGTGGTGTGCTGGGTGAGTGTGTAACACAGA 420
 DB CTCTAACGAAGTATTCTGTTCTTAGTGTGGTGTGCTGGGTGAGTGTGTAACACAGA 63866

QY 421 CGTCATCAAGGAGACAGACAGTATTTGCTAATATGAAAGTACAGATTAATTTACACTA 480
 DB CGTCATCAAGGAGACAGACAGTATTTGCTAATATGAAAGTACAGATTAATTTACACTA 63926

QY 481 CATTGTACATGAGTAATTCAACTGAATAAAGTGTCAAGGTAAAGCTTTTAAACGGTT 540
 DB CATTGTACATGAGTAATTCAACTGAATAAAGTGTCAAGGTAAAGCTTTTAAACGGTT 63986

QY 541 AATTTCCTGTCACAAACAGTAGATGACAAATGCGCGATCTTATCAGTGTCTCTCTTGAGCCCC 600
 DB AATTTCCTGTCACAAACAGTAGATGACAAATGCGCGATCTTATCAGTGTCTCTCTTGAGCCCC 64046

QY 601 CTTTCCCTCTGCTGTCCCTCCCAAGATGGGGGTGAGTCCATATTTAAACTGGCCATCC 660
 DB CTTTCCCTCTGCTGTCCCTCCCAAGATGGGGGTGAGTCCATATTTAAACTGGCCATCC 64106

QY 661 TCACAGTGTCTAACTTAGCAAGTCTTTCTTTAGACCCCTCTTAAAGCAACAATATG 720
 DB TCACAGTGTCTAACTTAGCAAGTCTTTCTTTAGACCCCTCTTAAAGCAACAATATG 64166

QY 721 TCTGACCTCTACTATAAAGATCTTTCTGATAATGCAATTCGGAGATTTTTTGTGATAGT 780
 DB TCTGACCTCTACTATAAAGATCTTTCTGATAATGCAATTCGGAGATTTTTTGTGATAGT 64226

QY 781 AGAAGTGGTTCCTGTTTTCACCTTCTTACTCAGCTAGTGTCCCTTCCCTTGGTTT 840
 DB AGAAGTGGTTCCTGTTTTCACCTTCTTACTCAGCTAGTGTCCCTTCCCTTGGTTT 64286

QY 841 CTAGTAACCTGGGTGAGAAATCAGGTGCTGGGCTTTACAGTCTTTTAAACTATTTAGAT 900
 DB CTAGTAACCTGGGTGAGAAATCAGGTGCTGGGCTTTACAGTCTTTTAAACTATTTAGAT 64346

QY 901 ATTCTGAAACATCACTGTCTTCCAGAGTACCAACACTGTCAATGATGATCCGCCCCC 960
 DB ATTCTGAAACATCACTGTCTTCCAGAGTACCAACACTGTCAATGATGATCCGCCCCC 64406

QY 961 CTCTAGACCTCACCACGGGACACATGCTTCGGTACCTTTGGTCTGTGAGGTTTC 1017
 DB CTCTAGACCTCACCACGGGACACATGCTTCGGTACCTTTGGTCTGTGAGGTTTC 64463

RESULT 4

BC051053
 LOCUS 2281 bp mRNA linear ROD 14-APR-2003
 DEFINITION Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.
 ACCESSION BC051053
 VERSION BC051053.1 GI:29835221
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 STRAUSBERG, R.
 REFERENCE 1 (bases 1 to 2281)
 AUTHORS Direct Submission
 TITLE Submitted (11-APR-2003) National Institutes of Health, Mammalian
 JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: mgc@nih.gov
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

REMARK
 COMMENT

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 108 Row: f Column: 9
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

FEATURES

Location/Qualifiers
 1..2281
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CZECH II"
 /db_xref="taxon:10090"
 /clone="MGC:58966 IMAGE:5038182"
 /tissue_type="mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV."
 /clone_lib="NCI CGAP_Lu29"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"
 197..775

CDS

/codon_start=1
 /product="Unknown (protein for MGC:58966)"
 /protein_id="AAH51053.1"
 /db_xref="GI:29835222"
 /translation="MQALKCVVGDGAVGKTCLLISYTNAPGSEYPTVFDNYSANV
 HVDKGNLGVDTAGDEYDRLEPSYQDVELICFSLVSPASFENVRKWPYEV
 HHCPTDILVGTIKLDRDITIEKLEKLTPTVQGLAMAKEIGAVKYLBSAL
 TORGLKTVFDEARVLCPPEVKRKKCLLL"
 BASE COUNT 568 a 569 c 504 g 640 t
 ORIGIN

Query Match 66.8%; Score 679.2; DB 10; Length 2281;

Best Local Similarity 86.1%; Pred. No. 6.7e-191;

Matches 876; Conservative 0; Mismatches 113; Indels 29; Gaps 10;

QY	1	CCCCATCTTCTGTCAGATTAAAGATTGCCAAATACCTTGTGAACCTAAGTTGGTGTG 60
Db	1019	CCCCATCTTCTGTCAGATTAAAGATTGCCAAATACCTTGTGAACCTAAGTTGTGTG 1078
QY	61	CTGAGAACAGCTAAGCACTAGCTGTGTGAGAGA--CTTTGTCTTAAAGAGCTGCAGC 117
Db	1079	CTGAGAACAGCTAAGCACTAAGCTTCTTGAGAGACTTGTGCTAAGAGACTGCAGC 1138
QY	118	TTCTGGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCCGCTGCACACAGCAGCC 177
Db	1139	TTCTGGAGTCAGGGGTGCAGACCCCTCCATAGTTCCAGACCCGCTGCACACAGCAGCC 1198
QY	178	TCCTTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGT 237
Db	1199	TCCTTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGT 1258
QY	238	AACCTTTGTACTGACGTCAC--GATGGGTGTAAACAGCTGCTGCTTTGATTTCATAGTG 294
Db	1259	AACCTTTGTACTGACGTCACAGTGATGAGTGTGCACCTCAGCTCTTGTGTTTCATAGTG 1318
QY	295	AGTTCTCTAAATACAGCTGACCGGTTCTGAGGCTTTGAACAGAACTCTGCCTCCG 354
Db	1319	AGTTCTCTAAAGAACACCGGCTAGCTTTTGTGAGACTTTTGAACAGAACTCTGCTCCG 1378
QY	355	TGTTGCTCTAAAGAGTATCTGCTCCTAGTGTGGGTGTGGGTGTGGGTGTGGTGTGGAA 414
Db	1379	TGTTGCTCTAAAGAGTATCTGCTCCTAGTGTGGGTGTGGGTGTGGGTGTGGTGTGGAA 1436
QY	415	ACACGAGCTCATCAAGAGGACAGACAGTATTTTGACTAATATGAGTACAGATTATTT 474
Db	1437	ACACGAGCTCATCAAGAGGACAGACAGTATTTTGACTAATATGAGTACAGATTATTT 1496
QY	475	ACACTACATTGTACATGAGTAATTCACATGAATAAAGTGTACCGGGTAAAGCTTTTAA 534
Db	1497	ACACTACATTGTACATGAGTAATTCACATGAATAAAGTGTACCGGGTAAAGCTTTTAA 1556

QY	535	ACGGTTAATTTCTGTCAAAACAGTAGATGACAAATAGGCGCATCTTATCAGTGTCTC----- 589
Db	1557	ACGGTTAATTTCTGTCAAAACAGTAGGTGACAAATAGGCGCAATCTTATCAGTGTCTCTTGAG 1616
QY	590	-----TCATTGAGCCCCCTTCCCTGCTGCTCCCTCCCGAGATGGGGCGTTGAGTCCA 642
Db	1617	CATCCCTTTCGGCGCCCCCACCCTCCACTGCGCTCCCTCCCGAGATGGGGCGTTGAGTCCA 1676
QY	643	TATTTAAACTGGCCATCTCTACAGTGTCTAACTTAGCAAGTGCCTTTCTTTAGGACCCCC 702
Db	1677	TATTTAAACTGGCCACCTCAGAGTTCTAACTTAGCAAGTGCCTTTCTTTA-GAACCCCC 1734
QY	703	TTCTTAAGCAGCAATATGCTGACCTGTACTATTAAGATCTTTCTGATAATGCAATTCG-GA 761
Db	1735	TTCTTAAGCAGCAATATGCTGACCTGTACTATTAAGATCTTTCTGATAACGACAGAGTTT 1794
QY	762	GATTTTGGTAGATAGTAGAAGTGGTCTCCTCTTTTTCACCTTCCCTTACTCAG-CTGA 820
Db	1795	TTTCTTTGGTAGCTCAGTAGAAGTGGTCTCCTCTTTTTCGCTTACTTTACTCAGAGCTA 1854
QY	821	CTAGTGTCTCCCTTCTGTTTCTAGTAACTGGGTGTAGAAATCAAGTGTCTGGGCTTTTACA 880
Db	1855	GTTAGTGTCTTCTAGTTTCTAGCAACTAGTGTGCAATCATGTGTGTGAGCTTTTACG 1914
QY	881	GTTTTTAACTATTTAGATATCTGAAAC--ATCAGTGTCTTCCAGAGTAGTACCAACT 938
Db	1915	GTTTTTAACTATTTAGATATCTTAAACTATGAACTTCTAAACAGAGTAGTACCAACT 1974
QY	939	GTCATGTGATTGAGCGCCCTCTAGACTCACCACGGGACACATGCTTCGGT 996
Db	1975	GTCATGTGACTAATGCTG---CCTTAGACCTCCCGACGTGGGACAGACGCTTCTCTGT 2029

RESULT 5

BC003828

LOCUS

BC003828

DEFINITION

Mus musculus RAS-related C3 botulinum substrate 1, mRNA (cdna clone

MGC:6235 IMAGE:3593957), complete cds.

ACCESSION

BC003828

VERSION

BC003828.1 GI:13277917

KEYWORDS

MGC

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 2319)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshikiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Aranson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, M.A.,

Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2319)

Strausberg, R.

Direct Submission

JOURNAL

Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil@stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 11 Row: e Column: 22.

FEATURES source

Location/Qualifiers

1..2319
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:6235 IMAGE:3593957"
 /tissue_type="Mammary tumor. Metallothionien-RGF alpha model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI CGAP_Maml"
 /lab_host="DH10B"
 /notes="Vector: pCMV-SPORT6"
 1..2319
 /gene="Rac1"
 /notes="synonym: D5Brttd559e"
 /db_xref="LocusID:19353"
 /db_xref="MGI:97845"
 198..776
 /codon_start=1
 /product="Rac1 protein"
 /protein_id="AAH03828.1"
 /db_xref="GI:13277918"
 /db_xref="LocusID:19353"
 /translation="MQAICVVVGDAVGKICLLISYTNAPFGEVPTVFDNYSANV
 MVDKPNVLGMDTAGEYDRLRPLUSYPTDVFLLICFLSPASENVRKRWYFVR
 HHCNPPIILVSTKLDRDDKDTIEKLKKLPTPTPOGLAWAKAIGVAKYLECSAL
 TQGLKTVFDEAIRVLCPPPVKRRKCLLL"
 603 a 573 c 503 g 640 t

BASE COUNT
ORIGIN

Query Match 66.8%; Score 677.6; DB 10; Length 2319;
 Best Local Similarity 85.3%; Pred. No. 2e-190;
 Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;
 QY 1 CCCCTATTCTGCTCAGATTAGAAATTCGCAAAATACCTGTGAACCTAACTGCGTTGTG 60
 DB 1020 CCCCCATTCTGCTCAGATTAAATATGCCAAATACCTTCTGAACCTAACTGCGTTGTG 1079
 QY 61 CTGAGAACACGTAAGCACTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGAGACTGCAGC 117
 DB 1080 CTGAGAACACCTAAGCACTAAACTCTCTTCGAGAGACTTCTGTTGCTAAGAGACCGCAGC 1139
 QY 118 TTTCTGGCTCAGGGTGCAGACCTCCGCTAGCTCCGAGACCGGTGCACACAGACAGCC 177
 DB 1140 TTTCTGGACGAGGGTGCAGACCTTCCAGAGTTCCAGACCGGTGCACACAGACAGCC 1199
 QY 178 TCCTTAATGACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGT 237
 DB 1200 TCCTTAATGACAGTTCGCAATGTAACGCACTGTAACTTATCAGCCCATGTTTCATTAAGT 1259
 QY 238 AACTTTGCTACTGACGTCAC---GATGGGTGTAACAGCTCTGCTCTTTGATTTTCATAGTG 294
 DB 1260 AACTTTGCTACTGACGTCACAGTATGAGTGTGACAGCTCAGCTCTTTGATTTTCATAGTG 1319

QY 295 ACTTCTCTAAATACAGCTGACCGCTTCTCAGGCTTTGAACAGAACTCTGGTCCCTG 354
 DB 1320 AGTTTCTAAAGACACCGGACTAGCTTTTCAGACTTTGAACAGAACTCTGGTCCCTG 1379
 QY 355 TGTTCCTCTAAAGAAATATTCTGTTCTCTAGTCTGGGTGTGCTGGGTGAGATGTTGAA 414
 DB 1380 TGTTCCTCTAATGAATATTCTGTTCTCTAGTCTGGGTGTGCTGGGTGGA--GTGTGAA 1437
 QY 415 ACACGACGTCATCAAGGAGACAGACAGTATTCTGACTTAATATGAAGTACAGATTATTT 474
 DB 1438 ACACGACATGATCAAGGAGACAGACAGTATTCTGACTAATATGAAGTACAGATTACTTT 1497
 QY 475 ACACATACATTGTACATGGAGTAATCAACTGAATAAAAGTGCACGGGTAAAGCTTTTAA 534
 DB 1498 ACATACATTGTACATGGAGTAATCAACTGAATAAAAGTGCACGGGTAAAGCTTTTAA 1557
 QY 535 ACGGTTAATTTCTGTCAAACAGTAGATGACAAATGCCGATCTTATCAGTGTCTCTTGTG 589
 DB 1558 ACGGTTAATTTCTGTCAAACAGTAGGTGACAAATGCCGATCTTATCAGTGTCTCTTGTG 1617
 QY 590 -----TCTTGAGCCCCCTTCCCTCTGCTGCTCCCTCCAGATGGGGGTGAGTCCA 642
 DB 1618 CATCCCTTCGCGCCCCCACTCCACTGCGCTCCCTCCAGATGGGGGTGAGTCCA 1677
 QY 643 TATTAAACTGGCCATCTCTCAGATTGCTAACTTAGCAAGTGTCTTTCTTAGGACCCCC 702
 DB 1678 TATTAACTGGCCATCTCTCAGATTGCTAACTTAGCAAGTGTCTTTCTTTA-GAACCCC 1735
 QY 703 TTCTTAACGAGCAATATGCTGACCTGTACTAATAGATCTTCTGATAATGCA-----TT 757
 DB 1736 TTCTTAACGAGCAATATGCTGACCTGTACTAATAGATCTTCTGATAACGAGAGATTT 1795
 QY 758 CGGAGATTTTGTGTAGTAGTAGAAGTGGCTTCTCTGTTTTCACCTCTCTTACTCAG- 816
 DB 1796 TTGTGTTTTTTTGTGTAGTCTAGTAGAAGTGGCTTCTCTGTTTTCGCTTACTTACTCAGA 1855
 QY 817 CTGACTAGTGTCTTCCCTTCGTTTCTAGTAACCTGGGTGTAGAAATACGCTGCGGCTT 876
 DB 1856 GCTAGTAGTGTCTTCCCTTCTAGTGTCTTCCAGCAATCATGTGTGACGCTT 1915
 QY 877 TACAGTTTTTAACTATTTTAGATATTCTGAACATCATCTGTCTTGCCAGAGTACCAACA 936
 DB 1916 TACGGTTTTTAACTATTTTAGATATTCTTAACTATGAACCTTCTAAACAGTACCAACA 1975
 QY 937 CTGCTATGATGATGCGCCGCCCTCTAGACCTCACCACGCGGACACATGCTTCGGGT 996
 DB 1976 CTGCTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2032

RESULT 6
 AC068493 269081 bp DNA linear HTG 15-MAY-2002
 LOCUS Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
 DEFINITION SEQUENCE, 26 unordered pieces.
 AC068493
 AC068493.10 GI:15148081
 VERSION HIG; HTGS PHASE1; HTGS DRAFT.
 KEYWORDS Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 269081)
 Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
 High Throughput Mouse Sequencing
 Unpublished
 2 (bases 1 to 269081)
 Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
 Direct Submission

JOURNAL

Submitted (03-MAY-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

COMMENT

On Aug 11, 2001 this sequence version replaced gi:14993654.

-----Genome Center

Center: Harvard Partners Genome Center

Center Code: HPGC

Web site: <http://www.hpcgg.org/Sequence/mouse.html>

Contact: hpgc@model.mgh.harvard.edu

-----Summary Statistics

Center project name: ABN

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 100%

*Consensus quality: 256302 at least Q20

*Consensus quality: 254124 at least Q30

*Consensus quality: 250276 at least Q40

Estimated insert size: agarose-FP - N/A

**Estimated insert size: 268581 - sum-of-contigs

Quality coverage: agarose-FP - N/A

Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 51730: contig of 51730 bp in length
* 51731 51750: gap of unknown length
* 51751 82423: contig of 30673 bp in length
* 82424 82444: gap of unknown length
* 82444 118158: contig of 35715 bp in length
* 118159 118178: gap of unknown length
* 118179 139441: contig of 21263 bp in length
* 139442 139461: gap of unknown length
* 139462 160638: contig of 21177 bp in length
* 160639 160658: gap of unknown length
* 160659 177611: contig of 16953 bp in length
* 177612 177631: gap of unknown length
* 177632 190991: contig of 13360 bp in length
* 190992 191011: gap of unknown length
* 191012 200128: contig of 9117 bp in length
* 200129 200148: gap of unknown length
* 200149 211882: contig of 11734 bp in length
* 211883 220773: contig of 8871 bp in length
* 220774 220794: gap of unknown length
* 220795 229643: contig of 8850 bp in length
* 229644 229664: gap of unknown length
* 229665 238339: contig of 8676 bp in length
* 238340 238359: gap of unknown length
* 238360 244868: contig of 6489 bp in length
* 244869 249725: gap of unknown length
* 249726 249745: contig of 4857 bp in length
* 249746 252523: gap of unknown length
* 252524 252544: contig of 2778 bp in length
* 252545 257520: gap of unknown length
* 257521 257541: gap of unknown length
* 257542 260396: contig of 2856 bp in length
* 260397 260416: gap of unknown length
* 260417 262709: contig of 2293 bp in length
* 262710 262729: gap of unknown length
* 262730 263934: contig of 1205 bp in length
* 263935 264196: gap of unknown length
* 264197 264217: contig of 242 bp in length
* 264218 264629: gap of unknown length
* 264630 264649: contig of 413 bp in length
* 264650 265608: gap of unknown length
* 265609 265628: gap of unknown length

```

```

* 265629 265935: contig of 1307 bp in length
* 265936 265955: gap of unknown length
* 265956 268335: contig of 1380 bp in length
* 268336 268355: gap of unknown length
* 268356 268549: contig of 194 bp in length
* 268550 268569: gap of unknown length
* 268570 269081: contig of 512 bp in length.

FEATURES             Location/Qualifiers
     source            1..269081
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /clone="RP23-76K1"
                     /sex="male"
     misc_feature       1..51730
                     /note="assembly_name:Contig138"
     misc_feature       51751..82423
                     /note="assembly_name:Contig137"
     misc_feature       82444..118158
                     /note="assembly_name:Contig136"
     misc_feature       118179..139441
                     /note="assembly_name:Contig135"
     misc_feature       139462..160638
                     /note="assembly_name:Contig134"
     misc_feature       160659..177611
                     /note="assembly_name:Contig133"
     misc_feature       177632..190991
                     /note="assembly_name:Contig132"
     misc_feature       191012..200128
                     /note="assembly_name:Contig131"
     misc_feature       200149..211882
                     /note="assembly_name:Contig130"
     misc_feature       211903..220773
                     /note="assembly_name:Contig129"
     misc_feature       220794..229643
                     /note="assembly_name:Contig128"
     misc_feature       229664..238339
                     /note="assembly_name:Contig127"
     misc_feature       238360..244848
                     /note="assembly_name:Contig126"
     misc_feature       244869..249725
                     /note="assembly_name:Contig125"
     misc_feature       249746..252523
                     /note="assembly_name:Contig124"
                     clone_end:SP6
                     vector_side:left
     misc_feature       252544..257520
                     /note="assembly_name:Contig123"
     misc_feature       257541..260396
                     /note="assembly_name:Contig122"
     misc_feature       260417..262709
                     /note="assembly_name:Contig121"
     misc_feature       262730..263934
                     /note="assembly_name:Contig120"
     misc_feature       263955..264196
                     /note="assembly_name:Contig119"
     misc_feature       264217..264629
                     /note="assembly_name:Contig118"
     misc_feature       264650..265608
                     /note="assembly_name:Contig117"
     misc_feature       265629..266935
                     /note="assembly_name:Contig116"
     misc_feature       266956..268335
                     /note="assembly_name:Contig115"
     misc_feature       268356..268549
                     /note="assembly_name:Contig114"
     misc_feature       268570..269081
                     /note="assembly_name:Contig113"
BASE COUNT      69726 a 66027 c 64672 g 67724 t      932 others
ORIGIN

```

Query Match

66.8%; Score 677.6; DB 2; Length 269081;

Best Local Similarity 85.3%; Pred. No. 3e-190;		Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;	
QY	1	CCCCATTCTTCTCAGATTAGAAATGCGAAATACCTTGTGAACATAAGTTGCGTTGTG	60
Db	62207	CCCCATTCTTCTCAGATTAGAAATGCGAAATACCTTGTGAACATAAGTTGCGTTGTG	62266
QY	61	CTGAGAACGTAAGCACTAAGCTGTGAGAG---CTTTGTCCTTTAAGAACTGCACG	117
Db	62267	CTGAGAACGTAAGCACTAAGCTGTGAGAG---CTTTGTCCTTTAAGAACTGCACG	62326
QY	118	TTCTGGGCTCAGGGTGAGACCCCTCCCTAGTCTCCAGACCGTGTGACACAGCAGCC	177
Db	62327	TTCTGGGCTCAGGGTGAGACCCCTCCAGGTTCCAGAGTCCAGACCGTGTGACACAGCAGCC	62386
QY	178	TCCTTAATGACACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGT	237
Db	62387	TCCTTAATGACACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGT	62446
QY	238	AACTTTGTACTGTACTGAC---GATGGGTGAACAGCTCTCTCTTGTATTTCAATG	294
Db	62447	AACTTTGTACTGTACTGAC---GATGGGTGAACAGCTCTCTCTTGTATTTCAATG	62506
QY	295	AGTTCTCTAAATATACAGCTGACCGGCTTCTGACGCTTTGAACAGAACTCTGGCTCCCTG	354
Db	62507	AGTTCTCTAAATATACAGCTGACCGGCTTCTGACGCTTTGAACAGAACTCTGGCTCCCTG	62566
QY	355	TGTTGCTCTAAAGAAATTTCTGTTCTTGTGCTGGGTGCTGCTGGGTGAGTGTGAA	414
Db	62567	TGTTGCTCTAAAGAAATTTCTGTTCTTGTGCTGGGTGCTGCTGGGTGAGTGTGAA	62624
QY	415	ACACGACGTCATCAAGAGGACAGACAGTATTTTGTACTTAATATCAAGTATGAGTTT	474
Db	62625	ACACGACGTCATCAAGAGGACAGACAGTATTTTGTACTTAATATCAAGTATGAGTTT	62684
QY	475	ACACTACATGTATACATGAGTAACTCAATGAATAAAGTGCACGGGTAAAGCTTTTAA	534
Db	62685	ACACTACATGTATACATGAGTAACTCAATGAATAAAGTGCACGGGTAAAGCTTTTAA	62744
QY	535	AGGTTAAATTTCTGTCACAGTATGATGACAAATGGCGGATCTTATCAGTGTCTC	589
Db	62745	AGGTTAAATTTCTGTCACAGTATGATGACAAATGGCGGATCTTATCAGTGTCTC	62804
QY	590	-----TCTTGAGCCCCCTCCCTCCCTGCTGCTCCCTCCAGATGGGGTGTGAGTCCA	642
Db	62805	CATCCCTTCGGCCCCCACCCTCCCTGCTGCTCCCTCCAGATGGGGTGTGAGTCCA	62864
QY	643	TATTTAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTGTCTTTCTTTAGGACCC	702
Db	62865	TATTTAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTGTCTTTCTTTAGGACCC	62922
QY	703	TTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTGATATGCA-----TT	757
Db	62923	TTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTGATATGCA-----TT	62982
QY	758	CGGAGATTTTGTGGTATAGTATGAGTGGTTCCTGTTTTCACCTCTCTTTACTCAG-	816
Db	62983	TTTGTGTTTTTGTGGTATAGTATGAGTGGTTCCTGTTTTCACCTCTCTTTACTCAG	63042
QY	817	CTGACTAGTGTCTTCTCTCTTCTAGTAACTGGGTGTAGAAATCAGCTGTGCGGCTT	876
Db	63043	GCTAGTGTGCTTCTCTCTTCTAGTAACTGGGTGTAGAAATCAGCTGTGCGGCTT	63102
QY	877	TACAGTTTTTAACTATTTAGATATTTCTGAACATCATCTGCTTCCGAGTACCAACA	936
Db	63103	TACAGTTTTTAACTATTTAGATATTTCTGAACATCATCTGCTTCCGAGTACCAACA	63162
QY	937	CTGCTATGCTAATGATGCGCCCTCTAGACCTCACCACGCGGACACATGCTTCCGCT	996
Db	63163	CTGCTATGCTAATGATGCG---CCTCTAGACTTCCGACGCTGAGACAGCGCTTCTGT	63219

AC102775
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC102775 185245 bp DNA linear HTG 23-MAR-2003
Mus musculus clone RP23-115C10, WORKING DRAFT SEQUENCE, 9 unordered pieces.
AC102775
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 185245)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-115C10
Unpublished
2 (bases 1 to 185245)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, I., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 185245)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, I., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Corum, B., DeArellano, K., Faro, S., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2003 this sequence version replaced gi:22381790.
All repeats were identified using RepeatMasker:
Smith, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L19647
 Center clone name: 115_C.10
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 183758 bases at least Q40
 Consensus quality: 184136 bases at least Q30
 Consensus quality: 184291 bases at least Q20
 Insert size: 178000; agarose-fp
 Quality coverage: 12.6 in Q20 bases; agarose-fp
 Quality coverage: 12.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 53693: contig of 53693 bp in length
 * 53694 53793: gap of 100 bp
 * 53794 56675: contig of 2882 bp in length
 * 56675 56776: gap of 100 bp
 * 56776 60589: contig of 3814 bp in length
 * 60589 60690: gap of 100 bp
 * 60690 67210: contig of 6521 bp in length
 * 67210 67211: gap of 100 bp
 * 67211 82585: contig of 15275 bp in length
 * 82585 82686: gap of 100 bp
 * 82686 103118: contig of 20433 bp in length
 * 103118 103219: gap of 100 bp
 * 103219 132145: contig of 28927 bp in length
 * 132145 132146: gap of 100 bp
 * 132146 182336: contig of 50090 bp in length
 * 182336 182436: gap of 100 bp
 * 182436 185245: contig of 2810 bp in length.
 * 185245 Location/Qualifiers
 * 1. 185245
 * /organism="Mus musculus"
 * /mol_type="genomic DNA"
 * /db_xref="taxon:10090"
 * /clone="RP23-115C10"
 * /clone_lib="RPCI-23 Female Mouse BAC"
 * 1. 53693
 * /note="assembly_fragment
 * clone_end:sp6
 * vector_side:left"
 * 53794..56675
 * /note="assembly_fragment"
 * 56776..60589
 * /note="assembly_fragment"
 * 60690..67210
 * /note="assembly_fragment"
 * 67311..82585
 * /note="assembly_fragment"
 * 82686..103118
 * /note="assembly_fragment"
 * 103219..132145
 * /note="assembly_fragment"
 * 132246..182335
 * /note="assembly_fragment"
 * 182436..185245
 * /note="assembly_fragment
 * clone_end:r7
 * vector_side:right"
 *
 * BASE COUNT 56870 a 34956 c 35069 g 57550 t 800 others
 * ORIGIN

Query Match 39.6%; Score 403.2; DB 2; Length 185245;
 Best Local Similarity 89.7%; Pred. No. 1.9e-108;
 Matches 481; Conservative 0; Mismatches 43; Indels 12; Gaps 4;
 QY 1 CCCCTATTCTTGTCTAGATTAAAGATTGCCAAATACCTTGTGAACCTAAAGTTGGCTTCTG 60
 Db 74 CCCCTATTCTTGTCTAGATTAAAGATTGCCAAATACCTTGTGAACCTAAAGTTGGCTTCTG 133
 QY 61 CTGAGAACACGTAAAGCACTAAGCTGTGTTGAGAGA---CTTTGTCTCTTAAGAAAGATGCAGC 117
 Db 134 CTGAGAACACCTTAAGCACTAAGCTCTCTTGAGAGACTTCTTGTCTTAAGAAAGATGCAGC 193
 QY 118 TTCTGGGCTCAGGGGTGAGACCCCTCCGCTAGCTCCAGAGCCGCTGTGACACAGACACGCC 177
 Db 194 TTCTGGAGTCTAGGGGTGAGACCCCTCCATAGTTCCTCCAGACAGTGTGACACAGACACGCC 253
 QY 178 TCCTTAATGACACGCTGCCATGTAAACCACTGTAACTTATCAGCCCATCTCATTCAGT 237
 Db 254 TCCTTAATGACACGCTTGCCATGTAAACCACTGTAACTTATCAGCCCATCTCATTCAGT 313
 QY 238 AACTTTCTACTGTACGTCAAC---GATGGGTGTAAACAGCTCTGCTCTTTGATTTCATATG 294
 Db 314 AACTTTCTACTGTACGTCAACGTGATGAGTGTGACAGCTCAGCTCTTTGATTTCATATG 373
 QY 295 AGTTCTCTAAATACCACTGACCGGCTTCTGAGGCTTTGAAACAGAACTCTGCTCCTG 354
 Db 374 AGTTCTCTAAAGACACCGGCTAGCTTTTGCAGACTTTTGAACAGAACTCTGCTCCTG 433
 QY 355 TGTTCCTCTTAAACGAAGTATTCTTCTAGTCTGTGGTGTGCTGGGTGAGTGTGTGAA 414
 Db 434 TGTTCCTCTTAAAGTATTCTTCTAGTGTGGTGTGCTGGGTGGA--GTGTGAA 491
 QY 415 ACAGCAGCTCATCAAGG---AGACAGACAGATTTTGTACTAATATGAAGTAGAGATTA 470
 Db 492 ACAGCAGCTCATCAAGGAGACAGACAGACAGATTTTGTACTAATATGAAGTAGAGATTA 551
 QY 471 ATTACTACTATGTACATGGAGTATTCACTGAATAAAGTGTCAACGGTAAA 526
 Db 552 CTTTACTACTATGTACATGGAGTATTCACTGAATAAAGTGTCAATGGGGAA 607
 RESULT 8
 AL583884 202342 bp DNA linear ROD 24-DEC-2002
 LOCUS Mouse DNA sequence from clone RP23-324B16 on chromosome 15,
 DEFINITION complete sequence.
 ACCSSION AL583884
 VERSION AL583884.20 GI:27368255
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 202342)
 Smith, M.
 Direct Submission
 Submitted (21-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 Cambridgehire, CB10 1SA, UK. E-mail enquiries:
 On Dec 23, 2002 this sequence version replaced gi:26788018.
 Sequence from the Mouse Genome Sequencing Consortium whole genome
 shotgun may have been used to confirm this sequence. Sequence data
 from the whole genome shotgun alone has only been used where it has
 a phred quality of at least 30.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the


```
QY 303 AAAATACCAGCTGACCGGCTTCTGAGGCTTTTGAACAGAACTCTGCGCTCCCTGTTGGCT 362
Df 56568 AAAAGACCGGAGTACTAGTTTGCAGAGTTTGAACAGATCTCTGTTCCCTGTTGGCT 56627
QY 363 CTAACGAAGTATTCCTAGTCTGAGTCTGCTGGTGTGCTGGTGGAGTGTGTGAACACGAGC 422
Df 56628 CTAATGAAGTATTCCTAGTCTGAGTCTGCTGGTGTGCTGGTGGAGTGTGAACACGAGC 56687
QY 423 TCATCAAGGAGACAGACAGTATTTTGAAGTATTAAGTATTAAGTATTAAGTATTAAGT 482
Df 56688 ATCAAGGAGACAGACAGTATTTTGAAGTATTAAGTATTAAGTATTAAGTATTAAGT 56742
QY 483 TTGTACATGGAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 526
Df 56743 TTGTACATGGAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 56786

RESULT 10
AX163752 1022 bp DNA linear PAT 22-JUN-2001
LOCUS
DEFINITION Sequence 16 from Patent WO0138579.
ACCESSION AX163752
VERSION AX163752.1 GI:14544858
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Gould-Rothberg, B.E., Dipippo, V.A., Ramsah, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
JOURNAL gene expression in liver
Patent: WO 0138579-A 16 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 273 a 208 c 205 g 336 t
ORIGIN

Query Match 28.6%; Score 290.6; DB 6; Length 1022;
Best Local Similarity 64.5%; Pred. No. 4.5e-75;
Matches 664, Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCATTCTTGTTCAGATTAAAGAGTTGCCAAATATACCTTCTGAACACACTGCAATGTT 57
Df 1 CCCCATTCTTGTTCAGATTAAAGAGTTGCCAAATATACCTTCTGAACACACTGCAATGTT 60
QY 58 GTGCTGAGACACGTAAGCACTAGCTGTTGAGAGACTTTGCTCTTAAGAGACTGAGC 117
Df 61 GTGCCGAGACACCGGACACTGAACTTTGCAAGACCTTCTGTTTGAAGAGACGGTAGC 120
QY 118 TTCTGGGCTCAGG-GGTGACAGACCTCCCGTAGC-----TCCGACGCTGTGACACA 169
Df 121 TTCTGAGTTAGAGGTGACAGACTTGTCTCTATGTATGTTCTCAGATGCGTAAGACA 180
QY 170 GCAAGCCTCTTAATGACAGCTGCCATGTAAAGCACTGTAACCTTATCAGCCCATGCT 229
Df 181 GAACAGCCTCCGGAATGAAGCTTGGCCATGCACTACCACTGAGTGTAGCAGCAGTGT 240
QY 230 CATTACCTAATTTGTACTGCTACGTCAGGATGGGTGTAAAGCTCTGCTCTTTGAATTC 289
Df 241 CCGACATACATTTGTAAGTCTGTAATGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 300
QY 290 TAGTGAGTTCTTAAATACACGCTGACGCTTCTGAGGCTTTGAACAGACTCTGGC 349
Df 301 TTGTGATTTCTATGAGGATTTTCTGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 358
QY 350 TCCTGTGTTGCCCTTAACGAAGTATTTCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGT 409
Df 359 -----CTATTCTTCTATGAGATTTCTGTT--TAGCTGTGGGTGTGCTGGGTGAGTGT 412
```

```
QY 410 GTGAACACGAGCTCATCAAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Df 413 GT-----GTGATCAAAAGGACAAAGACAGTATTTTGACAAATATCAAGTAGAG-- 460
QY 469 TAAATTACACTACATTTGACATGGAGTAA---TTCACTGAATAAAGTGTACAGGGTAA 525
Df 461 --ATTACACTACATTTGACAGGAATGAAGTGTACAGGGTAAAGTCTCTAAAGGTTA 518
QY 526 AGCTTTTAAACGGTTAAATTTCTGTCAACACAGTAGATCAAAATGGCCGATCTTATCAGT 585
Df 519 ATTCTGTCAATCGAGTAGATGATGAAGAAAGGTTGGTATTATCAGAAATGTTTCT 578
QY 586 TCTCTTTGAGCCGCCCTTCCCTGCTGCTCCCTCCAGATGGGCGTTGAGTCCATAT 645
Df 579 TAAGCTTTTCTCTTACACCTGCCATGCTCCCAATTTGGGCATTTAATTCATCT 638
QY 646 TTAACCTGGCCATCTCAGCTTCTAAGTCTAGAGTGGCTTTCTTTAGGACCCCTTC 705
Df 639 TTAACCTGGTGTCTGTTAGTCTAAGTCTAAGTCTTTTCTTTATAGAACCCCTTC 698
QY 706 TTAACGAGCAATATGCTGACCTCTACTATAAGATCTTTCTGAATAATCAATTCGAGATT 765
Df 699 TGACTGAGCAATATGCTT-CCTTCTATTAATAATCTTTCTGAATAATCAATTCGAGATT 757
QY 766 TTTTGTGATAGTAGAGTGGCTGCTGTTTCACTTCTTCACTTCTTCTTCTTCTTCTTCT 825
Df 758 TTTTGTGATAGTAGAGTGGCTGCTGTTTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCT 812
QY 826 GCTTCCCTCTGTTTCTAGTAACTGGGTGATGAATCACTGCTGCTGCGCTTTTACAGTTT 885
Df 813 GCTTCCCTCTGTTTCTAGTAACTGGGTGATGAATCACTGCTGCTGCGCTTTTACAGTTT 872
QY 886 TAAACTATTTTAGATA-----TTTGAAACATCACTGTCTTGCCAGAG 928
Df 873 TAAATAATTTTAGATAATCTTAACTATGAACCTTCTTAACTATCACTGTCTTGCCAGAT 932
QY 929 TACCAACACTGTCTAGTATGATGATGCGGCCCTCTAGACCTACCCACGCGGACACATG 988
Df 933 TACGACACTGTCACTTGACCACTACTG-ACCCTCTTTACCTGCGCCACGCGGACACAG 991
QY 989 CTTCGGGTA 997
Df 992 CCTCCTGTA 1000
```

```
RESULT 11
BC050687 2302 bp mRNA linear PRI 11-APR-2003
LOCUS
DEFINITION Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho
family, small GTP binding protein rac1), clone MGC:60264
IMAGE:6149377, mRNA, complete cds.
ACCESSION BC050687
VERSION BC050687.1 GI:29792301
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2302)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (08-APR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DPF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
```

Qy	1	CCCCATTCCTGCTCAGATTAAGATGTCGCCAAATACCTCTGTGAACATAAGTTGC---GTT	57
Db	1050	CCCCCATTCCTGTTTCAGATTAAAGAGTTGCCAAATACCTCTGGAACACACTGCAATTGTT	1109
Qy	58	GTGCTGAGAACACGCTAAAGCACTAAGCTGTTGAGAGACTTTTCTCCTTAAGAAAGATCGCAGC	117
Db	1110	GTGCGAGAACACCGAGCACTGAACCTTTGCAAGACCTTCGTCTTTGAGAGACGGTAGC	1169
Qy	118	TTCTGGGCTAAG--GGTGCAGACCCCTCCCGTAGC-----TCCAGACCCGTGTGCACACA	166
Db	1170	TTCTGCAGTTAGGAGGTGCACACACTTGCTCTCCTATGTAGTCTCTCAGATCGCTAAAGCA	1239
Qy	170	GCACAGCTCCTTTAATGACAGCTGCCCATGTAAAGCACCCTTAACCTATCAGCCCATGCT	229
Db	1230	GAACAGCCTCCCGAATGAAGCGTTTGCCCATTCACCTCCAGCTAGTTAGCAGCAGCTGTT	1289
Qy	230	CATTACGTAACCTTTGTACTGTACGTACAGATGGGTGTAAACAGCTCTGCTCTTTGATTCCA	289
Db	1290	CCCGACATACATTGTACTGTAATGGAGTAGAGGTAGCAGCTCAGCTCTTTGGATCAGTC	1349
Qy	290	TAGTGAGTTCTCTAAAATACCAGCTGACCGGCTTTCTGCAGCCTTTTGAAACAGAACTCTGGC	349
Db	1350	TTTGTGATTTTCATAGCGAGTTTCTGACCAGCTTTTTCGGGAGATTTTGAACAGAACTG--	1407
Qy	350	TCCTGTGTGTGCTCTAACGAAGATATCTGTTCTCTAGTCGTGGGTGTGCTGGGTGGAGTGT	409
Db	1408	----CTATTTCCTCTAAATGAAGAATTCGTGTT--TAGCTGTGGGTGTGCGGGTGGGGTGT	1461
Qy	410	GTGAACACGACGCTCATCAAGAGGACACAGACAGTATTTTGTAC--TAATATGAAGTAGAGAT	468
Db	1462	GT-----GTGATCAAGGACAAAGACAGTATTTTGCACAAATACGAAGTGGAG--	1505
Qy	469	TAATTTACACTACATTGTACATGGAGTAA---TTCACCTGAATAAAGTGTACCGGTAA	525


```

misc_feature
/note="alternative transcription start site"
1037
/gene="rac1"
/note="alternative transcription start site"
1053
/gene="rac1"
/note="alternative transcription start site"
1106..1175
/rpt_family="GC rich"
Join(1227..1261,13701..13772,18413..18530,22804..22860,
24268..24330,26010..26169,26458..26588)
/gene="rac1"
/note="alternative"
/codon_start=1
/product="Rac1b protein"
/protein_id="CAA10733.6"
/db_xref="GI:8574039"
/db_xref="SPTREMBL:O95501"
/translation="MQAICVVVGAGVGKTCLLSYTTNAPFGEYIPTVFNYGANV
MVDGKPVNLGLMDTAGEDYDLRLPLSPQTVGETYGGKDTISRGKDKPIADYELICFS
LVSPASFNVRKAWPEVRHHCNPPIILVGHKDLRDDKDTIEKLKKKLLPTIPYFQ
GUAMAKEIGAVYLECSALTQRLKTVFDEAIRAVLCPPPVKKRKKCLLL"
Join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MQAICVVVGAGVGKTCLLSYTTNAPFGEYIPTVFNYGANV
MVDGKPVNLGLMDTAGEDYDLRLPLSPQTVGETYGGKDTISRGKDKPIADYELICFS
LVSPASFNVRKAWPEVRHHCNPPIILVGHKDLRDDKDTIEKLKKKLLPTIPYFQGLAVK
TORGLKTVFDEAIRAVLCPPPVKKRKKCLLL"
<1227..1261
/gene="rac1"
/number=1
1262..13700
/gene="rac1"
/number=1
1390..1419
/rpt_family="GC rich"
1752..1793
/rpt_family="FLAM"
1794..2005
/rpt_family="AluJo"
2140..2171
/rpt_family="Simple_repeat"
2666..2972
/rpt_family="AluSx"
2980..3267
/rpt_family="AluSx"
3268..3289
/rpt_family="(T)n"
3802..4102
/rpt_family="AluJb"
4122..4386
/rpt_family="AluSx"
4538..4719
/rpt_family="LTR33A"
4754..5053
/rpt_family="AluSx"
5297..5594
/rpt_family="AluSx"
5655..5757
/rpt_family="AluSx"
5758..6062
/rpt_family="AluSx"
6063..6198
/rpt_family="AluSx"
6222..6401
/rpt_family="AluSg/x"
7162..7238
/rpt_family="LIMC5"
repeat_region
/rpt_family="Simple_repeat"
7348..7367
/rpt_family="Simple_repeat"
7368..7656
/rpt_family="AluSx"
7880..8183
/rpt_family="AluSx"
8484..8778
/rpt_family="AluSx"
9071..9365
/rpt_family="AluSx"
10279..10583
/rpt_family="AluSp"
10609..10864
/rpt_family="L2"
10970..11276
/rpt_family="AluJo"
11276..11594
/rpt_family="AluJb"
11744..11825
/rpt_family="L2"
11764..11987
/rpt_family="MIR"
12989..13307
/rpt_family="AluSx"
13701..13772
/gene="rac1"
/number=2
13773..18412
/gene="rac1"
/number=2
13954..14235
/rpt_family="AluSx"
14807..14831
/rpt_family="Simple_repeat"
14836..15128
/rpt_family="AluSx"
15408..15715
/rpt_family="AluSx"
15872..15896
/rpt_family="AT rich"
15902..16186
/rpt_family="AluSx"
16187..16314
/rpt_family="AluJo"
16413..16585
/rpt_family="Simple_repeat"
16586..16679
/rpt_family="FLAM_A"
16707..16810
/rpt_family="LIM4"
16811..17108
/rpt_family="AluSg"
17109..17345
/rpt_family="LIM4"
17377..17686
/rpt_family="AluSx"
17703..17729
/rpt_family="Simple_repeat"
18413..18530
/gene="rac1"
/number=3
18531..22803
/gene="rac1"
/number=3
19287..19436
/rpt_family="MIR"
21161..21184
/rpt_family="Simple_repeat"
21185..21351
/rpt_family="AluSp"
21365..21403
/rpt_family="Alu"
21364..21365
misc_difference

```


/gene="rac1"
 /notes="2345 nt insertion sequence in RPII 11-derived BAC
 Query Match 28.6%; Score 290.6; DB 9; Length 28567;
 Best Local Similarity 64.5%; Pred. No. 5.9e-75;
 Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY	1	CCCTATTCCTGTCAGATTAAAGATTGCCAATAACCTTGTAACCTGAAGTTGC---GTT	57
Dd	26871	CCCCCATTCCTGTTCAGATTAAAGATTGCCAATAACCTTCTGAACCTACACTGCATTGTT	26930
QY	58	GTCCTGAGAACCGTAAGCACTAAGCTGTGTGAGAGACTTTGGTCCCTTAAGAAGACTGCAGC	117
Dd	26931	GTCCGAGAAACACCGAGCACTGAACCTTTCGAAGACCTTCGTCCTTTGAGAAGACCGTAGC	26990
QY	118	TTCCTGGGCTCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCCAGACCCTGTGCACA	169
Dd	26991	TTCTGCAGTTAGAGGTGCAGACACTTGTCTCCTATGTAGTTCTCAGATGGGTHAAGCA	27050
QY	170	GCACAGCCTCCTTAATGACACGCTGCCATGTAAACGCACTGTAACCTTATCAGCCCATGCT	229
Dd	27051	GAACAGCCTCCCGAATGAACGCGTTGCCAATTGAACCTCACCACTGAGTTAGCAGCAGTGTT	27110
QY	230	CATTAGCTAACTTTGTACTGTACCTCACGATGGGTGTACAGCTCTGCTCTTTGATTTCA	289
Dd	27111	CCGCAATAACATTGTACTGTGAATGGAGTGAGCGTAGCAGCTCTTTGGATCAGTC	27170
QY	290	TAGTGAGTTCTCTAAAATACCAGCTCACCGGCTTCTGCAGGCTTTTGAAACGAACTCTGCG	349
Dd	27171	TTTGTGATTTATGCGAGTTTCTGACCACTTTTTCGGAGATTTTGAACAGAACTG--	27228
QY	350	TCCTGTGTTGCCCTTAACGAAGTATTCCTTCTCTAGTCGTGGGTGTCTGGGTGGAGTGT	409
Dd	27229	----CTATTTCCCTCTAATGAAGAAATTCCTGTT--TAGCTGTGGGTGTGCCGGTGGGGTGT	27282
QY	410	GTGMAACA CGAGTCATCAAAGGAGACACAGATATTTTGAC-TAATATGAGTAGAGAT	468
Dd	27283	GT-----GTGATCAAAGGACAAAGACAGATATTTTGACAAATACGAAGTGGAG--	27330
QY	469	TAATTTACACTCATTTGTACATGGAGTAA--TTCAACTGAATTAAGCTGCAGGTTAA	525
Dd	27331	--ATTTACACTCATTTGTACAGGAATGAAGGTGCACGGGTAAAACTCTAAAGGTTA	27388
QY	526	AGCTTTTTAACGGTTAAATTTCTGTCAACAGTAGATGACAAATGCCGATCTTATCAGTG	585
Dd	27389	ATTTCCTGTCAAAATGCACTAGATGATGAAAGAAAGGTGTGCTATTATCAGAAATGTTTCT	27448
QY	586	TCTCTCTTGAGCCCCCTTCCCGCTGCTGTCCCTCCCGAGATGGGGGCTTGAGTCCAATAT	645
Dd	27449	TAACTCTTTCCCTTCTTTACACCTGCCATGCTCCCAAAATGGGCAATTTAATTCATCT	27508
QY	646	TTAAACTGGCCATCCCTCACAGTTGCTTAACCTAGCAAGTGCCTTTTCTTTAGCACCCCTTC	705
Dd	27509	TTAACTGGTGTGTTCTGTGTAGTCGTACTTAGTAAGTGCTTTTCTTATGAACCCCTTC	27568
QY	706	TTAAGAGCAATATGTCGTGACCTGTACTATAGAATCTTTTCTGATAATGCATTCGGAGATT	765
Dd	27569	TGACTGAGCAATATGCCT-CCCTTGATTATAAACTCTTCTGATAATGCATTAGAAGTT	27627
QY	766	TTTTTTGGTAGATAGTAGAAGTCGGTTCCTGTTTTCACTTCCTTTACTCAGCTGACTAGT	825
Dd	27628	TTTTTTGCGATTAGTAAAAAGTCCTTTCCATGTTAC-----TTTATTCAGAGCTAATAAGT	27682
QY	826	GCTTCCCTTTCGTTTTCTAGTAACTGGGTGAGAAATCACGGTCTGCGGCTTTACAGTTTT	885
Dd	27683	GCTTTCCTTAGTTTTCTAGTAACTAGGTGTA AAAATCATGTGTGCGAGCTTTATAGTTTT	27742
QY	886	TAAACTATTTTAGATA-----TTCTGAAAACATCACCTGTCTGCCAGAG	928
Dd	27743	TAAAAATTTTAGATAAATCTTTAAACTATGAACTTCTTTTACATCACCTGTCTGCCAGAT	27802
QY	929	TACCAACATCTGCATGTGATTGATCGCGCCCTCTAGACCTCAACCCACGCGGACACATG	988

174959 GTCCGAGAACCCAGCACTGACTTTGCAAGACCTTCCTCTTGGAGAGCGTAGC 175018
118 TTCTGGGCTCAGG-GGTGCGACACCTCCCGTAGC-----TCCAGACCGGTGTACACA 169
175019 TTCTGCGATTAGAGGTGCAGACACTTCTCTCTATGTAGTCTCAGATCGTAAGACA 175078
170 GCACAGCTCTCTAATGACACGCTGCCATGTAAAGCAGCTGTAACTATCAGCCCATGCT 229
175079 GAACAGCTCCCGAATGAAGCGTTGCCAATCAAGCTCAGCTAGTGTAGCAGCAGTGT 175138
230 CATTACGTAATTTGTACTGTACGTACAGATGGGTGTAAAGCTCTGCTCTTTGATTCA 289
175139 CCCGACATAAATTTGTAATGAGTGTAGAGTGTAGCAGCTCAGCTCTTTGGATCAGTC 175198
290 TAGTGAGTCTCTAAAATACAGCTGACCGGCTTCTGAGCGCTTTGAACAGAACTCTGGC 349
175199 TTTGTGATTTTATAGCGAGTTTCTGACCACTTTTTCGGAGATTTTGAACAGAACTG-- 175256
350 TCCTGCTGTGCTCTAAGAGTATTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGT 409
175257 -----CTATTTCTCTAATGAAGAAATTCGTT--TAGCTGTGGGTGTGCCGGGTGGGTGT 175310
410 GTGAAACACGACGCTCATCAAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGT 468
175311 GT-----GTGATCAAGGACAAAGACAGTATTTTGACAAATACGAAGTGGAG-- 175358
469 TAATTTACACTATGTTGATGAGTAA--TTCACTGAATAAAGTGTACGGGTAA 525
175359 --ATTACACTATATTGTACAAAGAAATGAAGTGTACGGGTAAATACTCTAAAAGTTA 175416
526 AGCTTTTAAAGGTTAATTTCTGCAACAGTATGATGACAAATGCGCGATCTTATCAGT 585
175417 ATTTCTGCTCAATGCTAGATGATGAAGAAAGGTTGGTATTAATCAGGAAATTTTTCT 175476
586 TCTCTCTTTGAGCCCTTCCCTGCTGCTCCCTCCAGATGGGGGTGAGTCCATAT 645
175477 TAAGCTTTTCTTCTTCTACCTGCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 175536
646 TTAACATGGCCATCTCAGCTGCTAATCTAGCAAGTGTCTTTTCTTTAGGACCCCTTC 705
175537 TTAACATGGTGTCTGCTGCTGCTAATCTAGTGTGCTTTTCTTTATAGAACCCTTC 175596
706 TTAACGAGCAATATGCTGCTGCTGCTAATCTAGTGTGCTTTCTGATAATGCAATCGAGAT 765
175597 TGACTGAGCAATATGCTGCTGCTGCTAATCTAGTGTGCTTTTCTGATAATGCAATCGAGAT 175655
766 TTTTGTGATAGTAGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 825
175656 TTTTGTGATAGTAGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 175710
826 GCTTCCCTGCTTTCTGAGTAACTGGGTGTAGAAATCAGTGTGCTGCTGCTGCTGCTGCTGCT 885
175711 GCTTTCCTGATTTCTGAGTAACTGGGTGTAGAAATCAGTGTGCTGCTGCTGCTGCTGCTGCT 175770
886 TAAACTATTTTAGATA-----TTCTGAAACATCACTGCTTCCGAGAG 928
175771 TAAATATTTTAGATAATTTCTTAACTATGAACCTTCTTAACTATCACTGCTTCCGAGAT 175830
929 TACCACACTGTCTATGTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 988
175831 TACCGACACTGTCTATGACCAATCTG-ACCCTCTTTAACCCTGCTGCTGCTGCTGCTGCTGCT 988
989 CTTCGGGTA 997
175890 CTTCTCTGTA 175898

RESULT 15
AX408985
LOCUS 1232 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1632 from Patent WO0229103.
ACCESSION AX408985

AX408985.1 GI:21441690
Homo sapiens (human)
Eukaryota; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Alvares, C., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 1632 11-APR-2002;
GENE LOGIC INC (US)
Location/Qualifiers
1. 1232
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. D25274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN

Query Match 28.0%; Score 285.2; DB 6; Length 1232;
Best Local Similarity 64.4%; Pred. No. 1.9e-73;
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;
QY 8 TCTTGCTCAGATTAAAGATTGCCAAAATACCTTGTGTGAACCTAAGTTGC---GTTGTGTGA 64
Db 1 TCTTGTTTCAAGTAAAGAGTTGCCAAAATAGCTTCTGAACCTACACTGCAATGTTTGTGCCGA 60
QY 65 GAACACCTAAGCACTAAGCTGTTGAGAGACTTGTCTCTTAAGAGACTGCGAGCTTCTGGG 124
Db 61 GAACACCGAGCACTGAATCTTAGCAAGACCTTGTCTTTGAGAAGCGGTACTGTCTGCA 120
QY 125 CTCAGG-GGTGCGACACCTCCCGTAGC-----TCCCAGACCCGTGTGACACAGCAGC 176
Db 121 GTTAGGAGGTGCAGACACTTGTCTCTCTATGTAGTCTCAGATCGTAAAGCAGACAGC 180
QY 177 CTCCTTAATGACAGCTGCCATGTAAAGCAGCTGTAACTTATCAGCCCATGCTCATTAAG 236
Db 181 CTCGGAATGAAGCGTTGCCAATTAAGCTCAGCTGAGTGTAGCAGCAGCTGTTCCCGACA 240
QY 237 TAATTTGTACTGTACGTACAGTGGGTGTAAAGCTGCTGCTCTTGTATTCATAGTCAG 296
Db 241 TAACATTTGACTGTATGAGTGTAGCGTAGCTGCTGCTTGTGATCTAGTCTTTGTGA 300
QY 297 TTCTCTAAAATACAGCTGACCGGCTTCTGAGCGCTTTGAACAGAACTCTGGCTCTCTGT 356
Db 301 TTTTATAGCCAGTTTCTGACCACTTTTCGGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGCTCTTAAGAGATTTCTGTTCTCTAGTGTGGGTGTGCTGGTGGAGTGTGTGAAC 416
Db 355 TTTTCTCTTAATGAAGAAATCTGTT--TAGCTGTGGGTGTGCGGGTGGGTGTGT----- 407
QY 417 ACGAGCTCATCAAGAGAGACAGTATTTTGAC-TAATATGAAGTAGAGATTAATTTA 475
Db 408 -----GTGATCAAGAGACAAAGACAGTATTTTGACAAATACGAAGTGGAG----ATT 458
QY 476 CACTACATTTGATACAGAGT---AATTCAACTGAATAAAGTGTCAAGGTAAGAGCTTTT 532
Db 459 CACTACATTTGATCAAGGATGAAGTGTACGGGTAAAGAACTCTAAAGGTTTAAATTTCTG 518
QY 533 TAACGGTTAATTTCTGTCAAGAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTCT 592
Db 519 TCAATGTCAGTAGATGATGAAGAAAGGTTGTTATTTATCAGAAATGTTTCTTAAAGCTT 578
QY 593 TGAGCCCTCTCCCTCTGCTGCTCCCTCCAGATGGCGTGTGAGTCCATATTAAGT 652
Db 579 TTTCTTTCTTTACCTGCTGCTGCTCCCAATTTGGGCAATTAATCATCTTTTAAACT 638
QY 653 GGCCATCTCTCACAGTTGCTAACTTAGCAAGTCTTTTCTTTTAGACCCCTCTCTTAACGA 712
Db 639 GGTGTGTTCTGTAGTGTGCTAAGTGTGCTTTTCTTTATAGAACCCCTCTCTGACTGA 698
QY 713 GCAATATGCTGACCTGTACTATAGATCTTTCTGTGATAATGCAATTCGAGATTTTTTTGG 772

699	GCAATATGCGCT - CCTGTGATTATAAAATCTTTCTGATAATGATTAGAAAGTCTTTTTTGT	757
773	TAGATAGTAGAAGTGCCTGCTGTTTTCACCTTCCTTTACTCAGCTGACTAGTGCCTCC	832
758	CGATTAGTAAAGTGCCTTCCATGTTAC - - - - - TTTATTTCAGAGCTAATAAGTGCCTTCC	812
833	TTCTGTTTTCTAGTAAGTGGGTGTAGAAATCAGTGTGCGGCTTTACAGTCTTTTAAACTA	892
813	TTAGTTTTCTAGTAACTAGGTGTAAAAATCATGTGTGTCAGCTTTATAGTTTTTAAATA	872
893	TTTTTAGATA - - - - - TCTGAAACATCACATGCTTTGCGAGAGTACCCAC	935
873	TTTTTAGATAATCTTTAAACTATGAACCTCTTTAAATCAATCACTGCTTTGCCAGATTACCCAC	932
936	ACTGTGATGTGATTGATCGCCGCCCTCTAGACCTCAACCACGGGACACATGCTTCGGG	995
933	ACTGTCACTTGACCAATACTG - ACCCTCTTTACCTCGCCACGGGACACACGCTCCTG	991
996	TA	997
992	TA	993

Search completed: November 23, 2003, 14:02:01
Job time : 3817.98 secs